EXHIBIT A

REDACTED

Secure Data Interchange

Field of Invention

This invention relates to systems for the personalization of information delivery, including the delivery of advertisements, product information, news and features. The system of Secure Data Interchange provides users and vendors with absolute control over profile information, while enabling focused targeting of information, and profile interchange between different entities. The system provides the technical infrastructure for a market for profiles, evaluation, and personalized information and product delivery.

Problem

The introduction of cheap and powerful new information technology allows manufacturers, service providers, and stores (on-line and off-line) to collect information about the transactions and preferences of customers and users cheaply and efficiently. Moreover, new network connectivity enables different vendors to exchange profiles for common customers, either statically or dynamically, in order to build broad and detailed profiles across vendor domains. There exist many potentially powerful synergies between the data sets that are collected by different vendors and service providers, that can be leveraged to provide appropriate services and products to customers. When analyzed with the proper statistical tools these data sets can reveal fundamental patterns in the behavior of users, and enable a vendor to provide appropriate information to a user. Furthermore, access to user-profiles collected by other vendors can enable vendors to provide focused information delivery to first-time users, and also cross-market services with other appropriate vendors.

Electronic intermediaries that monitor the activities of users across different vendors and service providers can also collect data about the products and services that vendors provide. This data can be used, with appropriate analysis, to provide users with advice about relevant services and products. User profiles can be used to identify the goals, preferences and interests of users, vendor profiles can be used to relate the services and products provided by vendors to the profiles of relevant users. Users can benefit because they can find information more readily, and vendors can benefit because they can reach potential customers more easily.

The problem with the ease with which data can be collected, and the ability to readily integrate information that is collected in diverse transactions and activities, is that all of this information represents a significant challenge to the privacy of individuals. A visitor to a web site does not even have to buy anything for information about his activities to be monitored, recorded, and passed onto other web sites in future interactions - for example through the use of "Cookie" technology that is largely transparent to end users.

Thus, although there are many advantages to building large intra-and cross-industry databases, most companies by necessity must keep their data to themselves, and individual users must be on their guard for unrequested and inappropriate solicitations from vendors who misuse their personal information. What is required is a system that enables data exchange and analysis within a secure framework that ensures privacy and protects against misuse of personal and transactional information. There is a conflict between the privacy rights of customers and effective marketing, with a focus on using information gathered about customers to refine offers to users.

Both vendors and users can benefit from the exchange of some information on transactions and personal preferences, from users to vendors, and between vendors. In the same way that a large amount of information leads to information overloading, and makes it difficult for an individual to find the information that he/she is searching for-a dynamic marketplace with many vendors, products and services can make it difficult and expensive for a user to locate an appropriate product or service. Similarly, vendors would like to offer their services and products to users who are most likely to benefit, and most likely to make purchases. There would be clear and powerful synergies if technology existed to enable:

(a) secure evaluation of the value of data to a requesting party;
(b) secure data transfer in an environment that ensures that data privacy rules are protected at all times.

Vendors would find benefit in sharing data with other vendors; this would deepen their understanding of their customers' behaviors and preferences, especially if certain customers were traceable across several data sets. One could imagine an on-line music store sharing data with an on-line ticket dealer. Firstly, they would be able to augment their mailing lists with each other's customers. Secondly, they would increase their understanding of what kinds of music particular customers prefer. For example, by analyzing the particular types of music that individual customers prefer, the ticket vendor could target appropriate concerts to each customer.

Users (individuals) would benefit from sharing data with other users (individuals). This is already evident in the popularity of news groups and web pages catering to individuals with shared interests. By learning what other people with similar tastes and preferences have discovered and enjoyed, a user can sidestep information overload in the search for personally satisfying information.

Vendors and users can benefit from each other. An obvious example would be in the use of collaborative filtering for the marketing of targeted promotions; rather than being deluged with coupons and advertisements that are of absolutely no interest, a user would benefit by being presented with advertising that is highly relevant. In the process, the vendor would increase advertising response rates, boosting overall efficiency. Users could also benefit from the personalization of content at vendors' web pages, and well focused banner advertisements at other web sites that they visit.

While the above scenarios demonstrate the potential benefits of the sharing of data across different parties, they also present the possibility of the misuse of data: vendors could sell each other's private data to rivals, and users' information could be used against them. At present there is no technological solution that provides the many benefits of information exchange, as outlined above, but without the ability to misuse data.

This invention relates to a new technique for encoded storage and communication of information that allows aggregate information to be recovered while protecting the privacy of information pertaining to any one individual. This invention also relates to, but is not limited to, a particular system incorporating the above-mentioned technique in generating target profiles in a system for customized electronic identification of desirable objects. In one embodiment of this system a profile is maintained for each user that, in the general case, records both the user's demographic attributes and a record of the user's buying habits or other expressed preferences. However, the profile is maintained in an encoded form to be described. With this encoding it will be impossible to extract particular information pertaining to an individual user. It will only be possible to get information about large groups to which the user belongs.

In this system a user's profile is maintained by continuously acquiring information about past decisions by the user and other "similar" users. By encoding the feedback from the user using the new technique the particular decisions made by the user are not revealed to anyone else.

Information gathered about commercial or other activities by individual users can be used to provide a higher quality of service to such users in a number of applications such as electronic commerce and governmental services. Competitors and law enforcement agencies can also use such information. Thus it is important to allow access to such information only as permitted by the user or as permissible by law.

Solution

The above-described problems are solved, and a technical advance achieved, by the system of Secure Data Interchange. The Secure Data Interchange architecture and system enables the profiles of users across many vendor types to be combined into portfolios and system-level aggregate information to the extent that a user specifies, within a system that is secure to manipulation from vendors. SDI presents a technical solution that alleviates some of the tensions and conflicts that exist between privacy and focused information delivery. SDI allows vendors to personalize, while users can remain in absolute control of their personal profiling information.

Profile information is released to a central Secure Data Interchange data warehouse, or to data warehouses provided operated by vendors or third parties, and can be both anonymized and randomized to protect the privacy of individuals, while allowing detailed statistical analysis and model building for dynamic on-the-fly personalization of services. Secure Data Interchange resolves many of the conflicts that exist between the benefits of personalization and well-focused products and services, and protecting the privacy of users. In particular, privacy is protected through restrictions on the amount of explicit information that is released about users, controls over the mechanisms with which vendors can track users in between sessions, and protecting the release of implicit data (e.g. "clickstream" data, that is not transactional).

The primary application of SDI is to a Internet-based electronic commerce system, with individual customers "users" and larger vendors connected over a network of clients and servers. As such, SDI can be implemented as described with existing standards and protocols, including (but not limited to) the HyperText Transfer Protocol (HTTP) for client-server communication, extensible Markup Language (XML) for embedding meta data in Web documents, Java applets and Java script for client-side processing, and encryption methods, such as Secure Sockets Layer (SSL) and the X.509 standard for secure information transfer. Secure Data Interchange also complements and significantly extends current proposals for open profiling (Open Profiling Standard OPS), and user privacy controls (Platform for Privacy and Preferences), that are active projects of the World Wide Web consortium. The key advance provided within SDI is the ability to manage privacy and control personalization within an integrated system.

SDI allows users to specify privacy and data-release policies, and control the aggregation of information across different vendors. Furthermore SDI enables vendors to augment detailed transaction-based information with broader information about users that is collected from their extended interactions with other information and service providers, again within a privacy protected system. SDI provides a user with complete control over her identity as she browses the Internet and makes on-line purchases, placing the user in absolute control over what information each vendor can collect or receive about the user.

SDI can manage the information that a user wishes to release to vendors when he/she registers with a new system. A client-side SDI enabled proxy (or browser plug-in) controls information revelation, and provides support for pseudonymous and anonymous interactions. SDI has a distributed architecture, with accurate and complete profiling information about a user maintained on trusted clients, and randomized (and potentially anonymized) information pushed to centralized data warehouses, information that is still valuable for trend analysis and model-building.

The main technical solutions that are used within the system of SDI to enable privacy guarantees and personalization are: (1) Information-theoretic tools, such as releasing randomized profile information to third-parties and vendors, and removing identifying information from messages between users and vendors; (2) Distributed secure processing, such as locating user profile information on the client machine of the user, and using local secure processing to personalize generic information provided by vendors; (3) Cryptographic techniques for pseudonym generation and validation, and user authentication; (4) multilevel collaborative filtering techniques.

The Secure Data Interchange provides a number of primary functions. First, the system provides a secure environment where data can be collected, aggregated across a number of vendors and users, and analyzed such that the privacy and usage requirements of the user and vendor are protected. We are able to use technological solutions to provide guarantees about the information that is made available to vendors. Second, the system manages the privacy policy of a user, through the control of information and automatic management of pseudonyms. Third, the system allows vendors to provide personalization to users without accessing potentially sensitive and valuable profile information about users.

The system also provides support for electronic commerce functionality, such as providing a mailing service for targeted and authorized solicitations to a virtual and anonymous mailing list, and providing support for a system of certification and anonymous payments. The Interchange can also provide a privacy-protected market place for customer information, where certain types of data can be purchased, rented or sold by vendors in order to enable synergies to be realized between the data sets of independent service providers and vendors.

1. Definitions

For the purposes of this patent it is useful to separate the users of the Secure Data Interchange into two general classes: users and vendors.

A user is an individual (person), or possibly a group of individuals (people) that share similar interests, that transacts with multiple vendors, information providers, and service providers, and has an interest both in receiving personalized service and protecting his/her privacy with respect to those transactions. In particular, a user might require that two vendors with which he has had transactions cannot build an integrated profile on the basis of the user's independent transactions, with the goal of using information about one user-vendor session to provide a refined service in another user-vendor session. As another example, users might state particular types of vendors that are able to exchange information about the user, but for restricted purposes. A user is connected to the Internet through a dedicated host machine running client software, such as an Internet browser with an SDI plug-in or secure Java applet. Users include individuals shopping and browsing on the World Wide Web, whose extended purchasing and interaction behaviors can allow vendors to target their products and services.

A vendor may be an information provider, or a provider of goods and/or services, connected to users and SDI through a dedicated server and an SDI proxy server. Unlike users, vendors have less interest in pseudonymous interactions with different users, because vendors will typically have an interest in developing a strong identity. Like users, vendors may wish to exchange information on transactions with other vendors while maintaining tight control over the type of vendors that receive information, and also over the exact use that is made of the information. For example, a vendor might be willing to sell information on its transactions to another vendor that is not a direct competitor, but able to cross-sell related products. A vendor might also like to purchase information from users about the transactions that they have engaged in with competitors. Vendors include stores, such as grocery stores, car dealerships, on-line music stores, on-line bookstores, and also information providers, such as web portals and on-line newspapers. Vendors also include external organizations, such as credit agencies, HMOs, schools and police stations, who may hold information on individuals.

2. *Architectural Overview*

In this section we provide an overview of the invention, describing both its top-level architecture, and describing the key technologies that we use to support dynamic and powerful data synergies within a framework that explicitly protects the privacy of users. Figure 1 shows the top-level architecture.

The system of Secure Data Interchange is a technical solution to the problem presented by the conflicting goals of providing personalized information/products/solicitations to users (beneficial to users and vendors) and the rights/desires of individuals to privacy in transactions. In particular, a user might desire to control the vendors that have a lot of information about its preferences, and prevent certain vendors/entities from gaining information about particular transactions. A user might also like to be able to "keep control" of information that is potentially valuable to vendors and advertisers, and allow only restricted access to that information—and extract a payment for such access. This is the user-centric model of SDI. Similarly, vendors would like to be able to exchange profile/market-research type information with other vendors, but not competitors, and require a secure system for protecting information that they make available.

The main data structure within SDI is the profile. A profile is a record of pertinent information about a user, for example demographic information, information about web pages browsed, content information from web pages browsed, recent on-line transactions, etc. Within SDI we associate profiles with an OWNER, and sign them with a ALLOWED-USE key to indicate what the owner has restricted the use of profile information to. The main technique used to support privacy within SDI is a pseudonym, that can fully represent a user in all transactions without providing any information about a user's true identity.

Profiles may be located on client-machines, with users; at intermediate level SDI servers, for example at ISP servers; at vendor-level SDI servers; and also in a central SDI server. Users define privacy policies on their client machine, and the client machine ensures that all user-vendor interactions are consistent with a user's policy, and that they cannot be abused—even by a malicious vendor. Client-level proxies can monitor users can build profiles (for example, as users browse and interact with vendors), and vendor-level proxies can monitor users that interact with the same pseudonym in multiple sessions. Both client-level proxies and vendor-level proxies can provide profiles to the central SDI server, with associated conditions of use. The central SDI server can aggregate and validate new profiling information that is received. Client-level and vendor-level proxies can periodically request profile updates.

2.1 Main Modules

Users interact with the system of Secure Data Interchange through clients, that are general purpose computers with memory and a connection to a network of other computers (clients and servers), for example via the Internet. The User represents an individual that is interested in performing multiple on-line and off-line transactions, with multiple vendors and other users, within the managed data interchange and privacy framework of SDI. Each user physically interacts with the Internet via his/her host machine. The host machine need not be the same machine for all sessions, and could include a user's home PC, work PC, and also any portable devices that the user has configured for SDI. The host machine runs client software, which will include an Internet browser, and also the client-level SDI proxy server, or appropriate browser plug-ins to make the browser SDI-enabled. A host machine is in general a computer with a CPU,

disk storage, a network connection, and main memory. We assume that the user trusts the host machine and client software.

Vendors interact with the system of Secure Data Interchange through servers, that are general purpose computers with memory and a connection to a network of other computers (clients and servers), for example via the Internet.

In addition to clients and servers, there are SDI modules that run on the machines, termed "client-level proxies" and "vendor-level proxies". The client-level proxy is a core component that interprets user messages and makes sure that all interactions with vendors satisfy a user's privacy and data-use policies. The vendor-level proxy enables vendors to interact with other key SDI entities: client-level proxies and the central SDI database.

Another SDI module resides on a gateway between user's and the Internet, to protect the pseudonymity of a user further. For example, when the user's client machine resides on an intranet (eg that of an ISP) then there is an ISP-level proxy server at the gateway to the Internet that ensures no identifying information is provided to servers. The ISP-level proxy servers also support pseudonymous email addresses for users, and may maintain a database of profiles for user pseudonyms in their user-base.

The other main SDI module is a central SDI server, that maintains detailed records of profiles for every user's pseudonym. Having all the information available for access at a unified level allows extensive data mining and collaborative filtering techniques to be applied, but still without violating user privacy and data use policies. In fact, the profile information can be physically distributed, for example located on the SDI-level proxies for each user base.

For example, it is possible to provide recommendations to a user under one of his/her pseudonyms through collaborative filtering techniques, even without any other information about the user—just from similar profiles from other users. Similarly, it is possible to recommend new prospects to vendors from the user base of other vendors, without actually providing vendors with profile information about prospective customers. Furthermore, the system of SDI can ensure that vendors can only provide impressions to users if it is allowed within a user's privacy policy.

The function of the central SDI server is to make as much use of profile information that is collected (and authorized) by a user for a particular pseudonym, by analysis of the profile information. The goal is not to try to augment the profile with more data—for example by adding demographic information or purchasing transaction records from other vendors. This is impossible within SDI because pseudonyms cannot be associated with a real-world identity unless authorized by a user, and information provided in the profile for a pseudonym is carefully filtered to prevent user identification.

2.2 Overview of System

The client-level proxy manages all interactions that a user has with other users and vendors. Essentially, the proxy interprets a user's messages to other users and vendors, and makes sure that a user's privacy and data-user policies are followed. The proxy also maintains up-to-date profiles for users, and allows vendors to personalize information that they provide to users on the client machine, without receiving access to a user's profile. The proxy is also authorized to automate authentication to vendor servers, and release of certain types of information.

A key mechanism used within Secure Data Interchange is that of pseudonymous interactions. This allows users to maintain long-term relationships with vendors, and release personal information to vendors, without any other party being able to use the information—is a system of "pseudonyms". Essentially, pseudonyms allow a user to separate its real-life identity from its identity with another user, or a vendor. The client-level proxy is also careful not to provide a vendor with any information that would compromise a user's identity. Pseudonyms provide a very useful middle-ground between total anonymity and complete disclosure. Vendors can still keep useful records of their transactions with a single user, because the system maintains a persistent pseudonym for each user with the same vendor.

Users can however choose to interact with each vendor in the system under a different pseudonym, to protect their identity and prevent information transfer across vendors. The control of pseudonyms provides users with a method to control the exchange of profile information between vendors.

A pseudonym allows a user to maintain a number of persistent relationships with different vendors or groups of vendors, with complete assurance that the vendors cannot use the pseudonyms themselves to build a profile of the user from the user's sessions across different vendors. A pseudonym provides: Identification, Authentication, Encryption, and Contact information. Pseudonyms are triples: (Pseudonym ID, Private Key, Pseudonym e-mail address).

Personal information collected through transactions across multiple vendors by the same user, under different pseudonyms is protected from transfer between vendors. Even when vendors themselves maintain local records of the transactions that they have performed with users, and response to queries that they have sent to users, there is no way of providing information in a form that another vendor can combine with its own information and use to personalize its service to the same user. This is because no two pseudonyms can be linked except at the (trusted) client machine. In addition, "randomized aggregates" are used when providing information to a vendor that might allow two vendors to link their pseudonyms. For example, the zip code, age, hair color, and profession of a user might all be useful information for a vendor—but providing all that information accurately to more than one vendor could allow the identity of a user to be compromised and vendors to exchange information.

** DP comment we are also careful to support different keys for interactions between users and vendors -- c.f. new "light security protocol paper".

Pseudonymous profiles are still useful in the aggregate, as part of a collaborative filtering system, even when noise has been added to some fields. The system of SDI has a centralized database of profiles and pseudonyms, that can be used for data mining and other collaborative filtering techniques. For example, suppose a user has a pseudonym that he/she uses to interact with all on-line book stores. Then the centralized database can perform collaborative filtering using the profile associated with that pseudonym and other profiles in the database to make personalized recommendations, for example on behalf a vendor that pays to receive such a service.

A vendor that belongs to SDI can use the profile associated with a user's pseudonym to provide a personalized interaction with the user. In one version the profile is released to a vendor, and the vendor can push personalized information to the user. In another version, the vendor can push generic information to the user's client, for personalization on the client. It remains important that the vendor does not learn the true identity of a user if the user is to prevent a vendor from providing information on the user to other vendors that know the true identity of the user. The user's personal information is only safe when it is not possible to associate information with anything except for the pseudonym of a user, and only one vendor can interact with the pseudonym (through SDI).

We also allow the release of anonymous/pseudonymous profiles to vendors to allow a vendor to provide "in-house" collaborative filtering, without the ability to target any of the users. The pseudonyms are only useful within SDI. Furthermore, because a user's communication channels with vendors is controlled within SDI, we can provide vendors with "rights" to solicit users, and rented (non-transferable) solicitation rights.

In a vendor-centric model of SDI we allow vendors to provide information (e.g. profiles about their user-base) to other vendors within a framework that prevents competitors from gaining a competitive advantage. This is possible because one role of the central SDI database is to act as a clearinghouse for profile information, that matches vendors and suggests data synergies. The central SDI database will not release profile information to competitors, according to rules that are designated by the submitting vendor.

Furthermore, even when a user designates that a group of vendors can exchange information on the basis of user-vendor transactions, one variation provides each vendor with a unique pseudonym. The only way that vendors can take advantage of profile information from user interactions with other vendors is via the central SDI database, that has a link between the pseudonyms that a user has with each vendor.

The system of SDI can track a user's browsing behavior, by cooperation between the client and the ISP-level proxy. The client-level proxy releases sequences of URLs that a user browsers, associated with the pseudonym under-which a user surfs at any particular time.

2.3 Ancillary Systems

SDI provides ancillary systems that are necessary to support a pseudonymous e-commerce system, for example independent certifying authorities, that are able to validate "once-in-a-lifetime" pseudonyms, that allow a user to prove that it only has one pseudonym for all interactions with a vendor. Certifying authorities can also issue certify properties of the user that owns a pseudonym, for example his/her credit worthiness, his/her age, his/her nationality — without requiring that the user identify themselves to a vendor or another user, and without breaking a link between pseudonym and real-world ID (even with the Certifying authority itself). The issued patent <<Reference here>> includes a description of anonymous payment mechanisms and a secure certificate management system, that we incorporate in

this patent by reference. Users can gain new credentials through recognized certifying authorities under one pseudonym, and transfer the certificate to another pseudonym when transacting with another server.

Similarly, SDI supports a module that allows e-mail to be sent to pseudonyms.

SDI also places personal information (eg mailing address of user, credit card number, ...) etc with a trusted third party that has an agent-relationship with the user. When a vendor wants to (for example mail a physical item to a user) it receives certification from the TTP, and then provides that to another TTP with an agent-relationship with the user that mails the item to the mailing address, without the vendor itself ever receiving information about the mailing address.

Legacy/demographic information is integrated with pseudonymous profiles, even when SDI has no way to identify a user's real-life identity with a profile, through client-side updates. SDI provides clients with new information, and clients update profile information for pseudonyms with randomized versions of the new information as they choose.

Finally, within SDI there is a method to allow users to receive compensation, in the form of rebates and electronic cash, in return from revealing personal information to vendors (even information that is slightly randomized and cannot be shared with other vendors).

The technique of blinded signatures is used extensively within our system to reduce the amount of information that third parties, and SDI, have about users. For example, the technique allows a user to create publiclprivate key pairs without giving a certifying party knowledge about the key pair generated. No party within SDI is able to build a dossier that links the pseudonyms of users without explicit information provided by the user. The user has an absolute method to prevent vendors from generating combined user profiles. Blinded signatures are described by D Chaum (??), and incorporated here by reference. Originally developed for the purposes of anonymous digital cash, they readily extend to general certificates. It is not sufficient for a bank to sign a number, with a "$1" signature, because the bank knows the number it has signed —and can trace the cash. Digital signatures allow the bank to be "blinded" and then sign, and the recipient can remove the blinding factor from the signed number to receive a valid signed number. This ensures user privacy when spending digital cash.

2.4 Main Data Flows

TO CENTRAL SDI SERVER (not via ISP)

Figure 3 illustrates the basic data flows from the client-level SDI proxy and the vendor-level SDI proxy towards the central SDI data warehouse. The vendor-level SDI proxy only knows the pseudonym of a user when it executes sessions with a user and a particular pseudonym, along with what ever basic profile information and other explicit information the user provides the vendor's server with.

The client-level proxy server logs all web surfing activity that a user engages in under each pseudonym, but does not have all information about the interaction of each user within a session with a single vendor. The client-level proxy also knows the pseudonyms that are equivalent for a user, and can provide more integrated information than vendors.

FROM CENTRAL SDI SERVER (to client and vendor)

Figures 4 and 5 illustrate data flow from the SDI data warehouse server to the client-level and vendor-level proxy servers. The updated user profiles are periodically requested by client-level proxy servers (pull), to enable the users to maintain an up-to-date profile for each pseudonym that they operate. The update to the vendor-level SDI proxies occurs according to one of two main modes, as illustrated in Figure 6. A vendor may have sensitive information that it is not willing to release to the central SDI server. In this case, Figure 6 (b) the central SDI server releases profiling information to the vendor-level SDI proxy, where analysis is performed. In Figure 6 (a) the vendor releases all relevant information to the central SDI server, and receives updated profile information.

Figures 4 and 5 show the preferred mode of profile releases under our architecture. The Secure Data Interchange server releases pseudonym profiles to users, when verified requests are received from the proxy server that represents the pseudonyms. The SDI server also releases encrypted profiles to the proxy server, either during daily updates, or when a request for update is received from a proxy server. The proxy server maintains an encrypted profile for each pseudonym in its user-base. The profiles are encrypted with a private key of the SDI server. When the proxy server relays a message from a user to a vendor it also augments the message with this encrypted profile, so that the vendor can make use of profile information about the user if the vendor has purchased the analysis ability from SDI. SDI will supply analysis capabilities to vendors, that allow vendors to gain the benefit of analyzing profiles for new users, or

users currently in their user-base, but without violating the privacy rights of users-for vendors can access only the results of the analysis, not the profile itself.

2.5 Underlying Cryptographic Infrastructure

All messages sent between users and vendors can be encrypted to prevent anyone other than the intended recipient from being able to read them. There are many technical solutions to this problem, including asymmetric public key/private key schemes such as the RSA encryption technique (although, users would need a unique key pair for each pseudonym). New "light" security protocols, such as using asymmetric key cryptography for initial validation, followed by shared key encryption are attractive—especially when the asymmetric key infrastructure is inefficient when users need a different key pair for each pseudonym anyway. << cite the new paper from Bell labs here >>
We use cryptographic techniques to "digitally" sign messages, in order to validate information contained within a message. A digital signature is computed through encryption with a private key, known only to the certifier, but the signature can be verified with the corresponding public key. This provides a recipient with a high degree of confidence that the message was indeed generated as claimed. An example technology for generating signatures, or "message digests", is MD5.

3. The Client-level Proxy

The client-level proxy, implemented as a client program running on the user's client machine is responsible for managing all data transfer between the client (meaning the client machine and the user) and vendors, or other users. In particular, a key function of the client-level proxy is to implement a pseudonym-management policy for a user—that is able to exert complete control over the ability of vendors to compile data on users. The client-level proxy also negotiates privacy and data-use practices with vendor level proxies. The proxy also allows users to control the addition of demographic and other personal information to electronic profiles, and adds noise to data fields to protect user identity. Figure 2 shows the client-side view of the Secure Data Interchange.

The client-level proxy maintains profile information for a user's collection of pseudonyms, and allows the user to view and challenge profile information. The proxy also provides a rule-based interface to allow a user to select appropriate privacy/personalization policies.

The client-level proxy also retrieves pseudonymous e-mail for a user from pseudonymous e-mail boxes, and maintains shared keys for use with each vendor that the user interacts with. Finally, the proxy can personalized generic information provided by a server according to rules provided by a server and profile information stored at the client, and filter/validate incoming e-mail messages.

The proxy also provides ancillary services, such as automatic user-verification with web pages, a "secure cookie" system to allow servers to maintain stateful interactions with users without allowing the identity of users to be compromised by "flags" that are left on the client and retrieved by another vendor later.

The primary mechanism that protects the identity of a user across multiple vendors and service providers is the ability to interact pseudonymously with vendors. The user can choose a unique pseudonym for each third party with which helshe interacts, and be absolutely certain that helshe is the only party that knows his/her true identity, other than the trusted proxy server where the pseudonyms are registered. There is no way that a vendor can know anything about the transactions that a user has had with other vendors under alternate pseudonyms unless the user chooses to disclose the equivalence of pseudonyms, or use the same pseudonym across multiple vendors.

3.1 Initialization

The Client-level SDI proxy server runs on the user's client machine, and acts as an intermediary between the user and the Internet, intercepting all outgoing and incoming messages. Given that the user runs a standard Internet browser, e.g. Netscape or Internet Explorer, the proxy can be implemented as a plug-in got the browser, integrated directly into the browser, or downloaded as Java (or some-other platform-independent) code. The browser is configured to use the SDI

proxy as its proxy, and the SDI proxy itself connects through the ISP-level (or other intranet gateway) proxy server to the Internet.

A new user must down-load the client-level proxy server that will run on hislher local host machine (or an SDI-enabled browser), and configure hislher browser to connect to the first-level proxy server. Furthermore, the user must be located on an SDI-enabled network, where the Internet Service Provider has a SDI second-level proxy server at the gateway to the Internet. A new user connects to the main Secure Data Interchange through hislher browser (configured through the first-level proxy server), by entering the SDI URL (e.g. http://www.sdi.com). The log-in page will prompt for a new user-name and password.

3.2 New-user Registration

The client proceeds to automatically generate a unique SDI user ID code, and provide information about the user to a central SDI database -- although this information will not be linked to a user's pseudonyms. A flow-chart for the process of registering a new user with SDI is show in Figure 10.

When a user first registers with SDI the user provides the client-level proxy with personal information, such as its name, mailing address, and e-mail address (at a minimum). The client-level proxy registers the user with the central SDI server, providing the server with the name, address and e-mail address of the user. Other basic user information could include demographic information, for example a users job, marital status etc. The user can configure hislher client-level SDI proxy to release some of this information automatically to vendors.

At this stage the central SDI server must verify the identity of the user, and also check that the user is not already registered with SDI. The method for verifying the identity of a user could include requesting that the user provides his/her social security number, or some other institutional solution that is used for this purpose. In the future we could envisage an electronic system for such an identity procedure, but the method might require for the user to execute this initial step in person with the presentation of a recognized photo ID. The central SDI user ID server maintains a database of all users that are registered with SDI, and checks that that the user is not already registered with the system of secure data interchange.

When a new user registers with SDI the user must create a unique public keylprivate key pair. This key pair can be generated only once for a person, and although the central SDI user ID server does not know the key pair, the server can verify that a key pair is only generated once - because a new user must present proof of identity to establish an account. The unique key pair is used by the client-level SDI proxies to generate new pseudonyms for users, and to verify (when necessary) that a user has only one persistent pseudonym for each vendor.

The client-level proxy now generates a unique user identifier, UUID. This is blinded, and signed by the central SDI server so long as the identity of the user can be validated. The technique of blinded signatures is discussed in the Appendix. The client-level proxy now removes the blinding factor, and has a signed UUID that it uses when it is necessary to generate new pseudonyms and request new certificates. The central SDI server that validates new users has a public/private key pair for the purpose of validation, eg. (PKSDI, SKSDI).

The central SDI proxy also provides the user with a signed certificate of some universal identifier, such as its Social Security Number, that the user can use to generate other certificates from "User Certifying Authorities".

Furthermore, the UUID acts as a public-key, and the client-level proxy also generated a private-key. The client-level proxy can now sign messages with its private-key, and provide the signed to UUID, to verify that (1) the UUID represents a validated user; (2) it is the client-level proxy authorized to act for the user, because it has the private-key associated with the UUID. The technique of public key/private key cryptography is discussed in the Appendix. The client-level SDI proxy uses the private key to authenticate messages that it sends to other modules within SDI, such as Pseudonym administering servers.

The unique user ID for a user does not carry any information about the user, its sole purpose is to provide a unique identity.

The central SDI server does not know the unique user ID that is associated with a user because we use the method of blinded signatures, but nevertheless a third party or another SDI module receives a guarantee that the user ID has been validated by the central SDI user ID server, and that only one user ID was certified for the user. This is a useful

property because the system of SDI is absolutely and unconditionally secure from security violations that could lead to the security of a users identity to be compromised.

3.3 Selecting Privacy and Profile Management Policies

The next stage in registering a new user with SDI is to establish privacy and profile management policies. We describe a "pseudonym management policy", and a "profile management policy". A user can define how he/she wishes to interact with various classes of vendors (depending on the nature of the business that the vendor is engaged in), the kinds of uses to which the transactional information that a vendor collects can be put to, and the amount of information that a vendor is authorized to release. The user can also specify a "basic profile" that the user is willing to release to any vendor, irrespective of the vendor's policies. The profile-management policy is further broken down into the "data-release" policy and the "use-of-data" policy. The client-level proxy manages a user's interactions with vendors, to keep them within desired policies.

3.31 Pseudonym Management Policies

Pseudonym management, coupled with the ability to add noise to information provided, allows users to exercise complete control over the ability of vendors to collect and exchange information about them. The client-level proxy contains a rule-based interface that allows the user to define a pseudonym-management policy.

Abstract Policy Hierarchy:

Level 0 (Highest)

At the highest level of privacy a user chooses to interact anonymously with every vendor, so that vendors cannot even personalize its service to a user over an extended interaction, because the vendor will never know who the user is. An anonymous pseudonym is simply a one-time traditional pseudonym, where a PAS does not need to check that a user does not already have a pseudonym for a vendor.

Level 1

A user can choose to interact with every on-line server under a unique pseudonym. This completely prevents a vendor from knowing anything more about the user than the information that can be inferred from the interaction itself. So long as that information does not identify the user, a vendor cannot access any information from external databases (such as demographic information), and a vendor cannot access any information collected by other vendors about the user. The vendor can still personalize information that is displayed to the user, but only on the basis of its own historical information for that user.

Level 2

** DP—note that the mode of information release by vendors and users allows users to provide profile information to competitors, but only when the client-side SDI proxy can do a useful job of monitoring details of a transaction (in general this is hard).

At level 2 a user can choose to share pseudonyms for groups of vendors (for example a user might choose the same identity for all vendors that sell music and books), and receive personalized service from each vendor on the basis of her extended profile.

The system of SDI allows users to maintain "ownership" of the dossier of information by: (1) providing a different pseudonym ta each vendor, but providing the central SDI database with information about which pseudonyms are equivalent; (2) allowing vendors to take advantage of the combined profile when providing information/products to a user by allowing personalization on the client machine but without release of the complete profile for the group of vendors to a vendor.

** DP: The client proxy server will release a user's profile (for a pseudonym) to the central SDI server, but anonymously, so that the central SDI server can perform pseudonym level profiling in order to enhance vendor user models, without compromising the privacy of the user.—What about anonymous release of information to the central SDI server? Not sure how this works....

Level 3

At level 3 the user can interact with every vendor under the same pseudonym, or anonymously with some vendors until enough trust has been established.

Level 4

At the lowest level of privacy protection (none) the user can simply use a unique ID that is linked to his/her true identity to interact with every vendor. At this level the user has no control over the dossiers of information that can be collected by groups of cooperating vendors.

Figure 6 illustrates level 2 - which is the preferred mode of interaction. The client-level SDI proxy for a user submits a request for information to a vendor server, including the pseudonym of the user. The vendor's server can perform some personalization at the server level, based on data that it has accumulated from previous interactions with the user's pseudonym, and also push generic information and rules for processing the information to the client machine, where information is processed according to a user's extended profile for the group of vendors to which the vendor belongs. The rules could be implemented as Java code or Javascript, and the generic information pushed as XML documents, for example.

3.32 Implementation Details

The central Secure Data Interchange server categorizes vendors that register with SDI, by assigning labels from a set of classifiers, that indicates the business of the vendor (i.e. the services and products that a vendor provides). The set of classifiers might include: music goods, news media, vacation packages, groceries, clothes. We denote the classifiers abstractly as L1, L2, ., LN. There is also a label for vendors that are not currently registered with SDI, denoted L0. Given the labels, each vendor has an associated set of labels, denoted L(Vj), for example L(V4) = {L1, L4, L12}.

The user is able to configure an appropriate pseudonym-management policy at the first-level proxy level that runs on his/her host machine. The user assigns a pseudonym-management action to different vendor-classes. The set of actions that SDI provides can include, but are not limited to:

A. Anonymous interaction with vendors in this class (i.e. use a different pseudonym every time the user enters the Vendors' sites.)
B. One pseudonym per vendor in this class, but the same pseudonym for all visits to the same vendor.
C. One pseudonym for all vendors in this class, and the same pseudonym at all times.

Pseudonym-management action A provides stronger data-privacy than action B, than action C, but action C provides the most opportunity to vendors to exchange information on the user and provide personalized and informed service to the user. This is a basic tradeoff that the user must make - between privacy and personalization.

A policy maps a vendor to a "management action", i.e. with Vendor V I will use the same persistent pseudonym for all interactions, but it will be unique to that vendor.

The vendor's are grouped according to their "business classifier" labels. For example, group 1 might contain all business that sell books or CDs, group 2 all businesses that sell computer hardware, etc. Each vendor group has an associated management action, e.g. vendors in group 1 (books/CDs) can share the same pseudonym for the user—and exchange information as provided by the user and the vendors to the central SDI server. Other groups could be defined by the labels that vendors DO NOT have, e.g. group 3 could be vendors that do not sell credit cards.

Clearly, it is possible for a vendor to be categorized into more than one group, within these rules. For example a vendor that sells books and Compact Disks would be placed in the group of vendors that sells books, and a group of vendors

that sells CDs. When this occurs the rule-base must choose the most appropriate group, for example on the basis of profiling a group of vendors according to the profiles of their user-base. Alternatively, it is possible to "partition" an individual business into different core businesses that fit cleanly into a single group, and then ensure that a user's interactions respect the current group that the vendor is a representative of.

The user can choose a mapping from labels to groups that best fits his/her own privacy needs, and then continue by assigning pseudonym management actions to each group. For example, the user might decide that all vendors that sell flights, books, and music should be assigned the same pseudonym for all transactions (i.e. pseudonym action C), while all vendors that sell financing, loans, and credit cards should be assigned to the same pseudonym for all transactions (i.e. pseudonym action C), but to a different pseudonym than the first class of vendors. Similarly, a user might require a single pseudonym for each vendor that is not classified by SDI (pseudonym action B), and anonymous pseudonyms for vendors that are classified as suppliers of adult material.

We allow vendors to provide different service levels, depending on the pseudonym action that a user chooses. For example, a vendor might desire to provide a better level of service to a user that interacts with a persistent pseudonym than a user that interacts with an anonymous pseudonym. The first-level proxy sends a token that certifies the type of pseudonym-action policy that a user has chosen when a user connects to a vendor's site. This is explained in more detail in Section << add section number here >> below.

3.4 Profile management policy

The basic mechanism that a user has to prevent vendors exchanging information with other vendors about hislher transactions is the pseudonym-management policy. However, there is still a chance that a user's identity can "leak" to another vendor if the user has performed a transaction or provided information that reveals the user's identity. For example, if a user purchases a flight from Philadelphia to San Francisco on UB 004 that arrives in San Francisco at 11.50am on 3110199 from one vendor, and then immediately purchases a car rental from San Francisco International, indicating arrival on flight UB 004 from another vendor, then there is a high probability that it is the same user. If the first vendor notifies all major vendors that offer car rentals that it has just completed a sale of a flight ticket to a user with pseudonym C1, and another vendor sells an appropriate car rental to user with pseudonym C121, then the two vendors now have some positive evidence that the two pseudonyms might belong to the same user, and in future can try to cross-sell appropriate products.

3.41 VENDOR-SIDE PROFILE MANAGEMENT

This is where the role of the "profile management policy" is important. A user can assign a "data-release" action to each class of vendor, depending on the types of actions that a user does with a vendor. There is no way to prevent a vendor releasing the details of a transaction with a user to other vendors, and if the details allow other vendors to identify the user, then the users privacy can be compromised. However, we can control the type of information that a vendor provides to the central SDI server, for exchange with other vendors.

To help with this process, the central SDI server classifies the type of service/goods that a vendor provides according to whether they are "anonymous" high-volume goods such as compact disks or newspaper articles, or "non-anonymous", low-volume, possibly personalized goods, such as cars, flights, or property. A vendor must initially enter a contract with a user about the type of information that it can release to other parties, i.e. the choice of data- release actions can include, but is not limited to:

A. Release no information.
B. Release randomized information.
C. Release all information.

The vendor-level SDI proxy is able to randomize transaction information automatically if option (B) is selected, to allow the vendor to submit profile information to the central SDI server without breaking the user's privacy requirements.

The vendor must provide the central SDI server with its certificate of agreement with the user, when it submits information. All profile information can be verified by the user, and in the randomized section we also describe a technique to validate randomization of profile information. The client-level proxy maintains profile information for each of a user's pseudonyms, and allows a user to view and modify (challenge) profile information that has been provided by vendors and other parties. The proxy requests periodic profile updates from the central SDI server, providing validation via the PID and associated private key that the proxy is authorized to receive profile information.

For example, a user might choose that all vendors that sell flights can only release randomized information to the central SDI server or other vendors, while vendors that sell compact disks can release complete information on transactions to the central SDI server and other vendors (although this information will only relate to the actions of the user under the same pseudonym).

3.42 CLIENT-SIDE PROFILE MANAGEMENT

Orthogonal to pseudonym management is profile management, which determines how a client-level proxy will release profile information to vendors.

Level 0 (Highest)

Release no information.

Level 1

Release randomized information. The information is randomized before release to prevent any compromise of the user's identity from profile information that is too specific, while enabling useful personalization at the server level, and enabling useful analysis at the central SDI server.

Level 2

Release non-randomized profile information for a user under the relevant user pseudonym.

We ensure the accuracy and content of profile, given that updates can be made by any third party that is privy to information by allowing users (individuals) to specify privacy constraints that vendors must uphold in reporting information to other vendors and/or SDI

3.43 CLICKSTREAM DATA (CLIENT-SIDE POLICY)

Clickstream data must be logged at the client, because of proxy caching. This data is released periodically to servers. The client-level proxy server that runs on a user's host machine is in a unique position of being able to monitor the user across different pseudonyms and across different vendors' sites. For example the proxy-server can monitor:

1. Clickstream data across different vendors' sites and pseudonyms.
2. Data that is displayed to the user (text, names of graphics objects.)
3. Input provided by the user at the keyboard of the host machine.

We term all of this data "clickstream" data because it is gathered by passively observing the actions of the user, and not by direct question-and-response. The clickstream data policies that are available to a user can include, but are not limited to:

A. Release no information.
B. Only release data on the URLs of the most recent sites visited.
C. Release data about the URLs of the most recent sites visited, and the information displayed to the user.
D. Release data about the URLs, the information displayed, and the information entered by the user.

In addition, the clickstream data policy restricts the depth of information that is provided, i.e. for how many previous sites, and can also restrict the data released to data that was collected under the current pseudonym of the user. The data

that is released can also be randomized in the same way as explicit data is randomized, for example by removing time-stamp information.

3.5 Certificate Management

User certifying authorities issue certificates that certify properties about the user that is represented under a particular pseudonym. This service is important to a pseudonymous electronic commerce environment because vendors will require guarantees about certain properties of users that they do business with, such as the age and credit worthiness of a user. Although outside of the basic SDI framework, the existence of such authorities is assumed in the description of the basic operation of SDI.

When a user needs certificate C(P) for pseudonym P, a user can use the following steps, to gain a certificate without providing a User Certifying Authority with any new information. The user sends a message to the Certifying authority, containing its SSN certificate, that it received on initial registration with SDI. The User certifying authority can then check the certificate, and if it is valid and the user can be certified for the requested property, signs a blinded message to certify the new property—for a one-time pseudonym specially generated by the user. This certificate can then be unblinded, transferred to certify other pseudonyms that the user holds, and used for certification purposes << See D Chaum's work on digital certificates >>

Certifying Authority servers have key pairs, for example (PKCA,k, SKCA,k), which again are using for certification and encryption purposes.

The certificate can be trusted so long as the certifying authority keeps its private key secure. To give an example, a trusted third party can exist to certify that a user represented with a particular pseudonym is above 18 years old. The third party can maintain a private key/public key pair (SK_18, PK_18) and sign a message that includes the pseudonym of the user with its private key to generate a certificate, C_18(U), that will assure other parties that the user is 18 years old. This certificate can then be requested by other vendors and information providers, and checked for validity within a public key infrastructure that maintains a faithful copy of the public key of the certifying trusted third party.

The first-level proxy server is responsible for certificate management. When a certificate is required by a vendor, then the proxy server checks whether the certificate has been issued to the user under one its pseudonyms. If this is the case, then the proxy server will simply transfer the certificate to the appropriate pseudonym for the user with the current vendor, using a technique taught by D. Chaum << more information here >> There are various certifying authorities within the Secure Data Interchange system, that are able to use a verified social security number to provide certification about the user. When a new user registers with SDI, the central SDI server provides the following social security number certificate to the user, S((SSN, PKSSNC), SKSDI), that links the social security number of the user to a public key that was established for this purpose.

When the user needs a new certificate, the first-level proxy server creates a new key pair, using the method described above to certify that this is the first certificate of this kind that the user has requested. The first-level proxy server receives a validated new public key, S(PKP, SKCAS), that it sends to the Certifying authority. The first-level proxy server now generates a blinded certificate number, B(CERTP), and transfers the social security certificate that is signed by the central SDI server to its new pseudonym. The first-level proxy server forms the message M=(B(CERTP), S((SSN, PKP), SKSDI)), signs it with its secret key SKP, and sends it to the Certifying authority.

The certifying authority now verifies the public key, PKP, and then verifies that the user has the appropriate secret key, i.e. checks that the message M is correctly signed. The certifying authority continues by verifying that the user with pseudonym P has social security number SSN. Then the authority determines whether the user with social security number SSN has the correct property, and if it does, signs an association between the blinded certificate number that was provided by the first-level proxy server of the user and the public key of the user's pseudonym.

Finally, the first-level proxy server can remove the blinding factor, and it is left with a signed certificate that associate the public key for its new pseudonym, PKP, with the certificate number. The certificate is represented CP = S((CERTP, PKP), SKCA).

The certifying authority now knows that the user with pseudonym PKP has been issued with a certificate, and furthermore, the authority knows the social security number of that is related to the pseudonym. However, the user will never need to use the pseudonym again, because it can transfer certificates issued in one pseudonym to certificates issued in another valid pseudonym using a method taught by D. Chaum << more details here >>. The Certifying authority therefore gained no information, because it already knew the relation between social security number and property that it is certifying. For example, we can transfer certificate CP to certificate CQ = S((CERTP, PKQ), SKCA).

<< actually, we need to also change the certificate number, else other vendors can still form a portfolio >>

3.51 Pseudonymous Interaction with a Vendor

The proxy can validate that it represents a particular user under pseudonym P by sending a message to a vendor with the signed PID, and signing a challenge provided by the vendor. The vendor can validate that the signature corresponds to the PID. See the Appendix for a discussion of this challenge-response mechanism.

3.6 Generating a New Pseudonym

The fundamental model of interaction between a user and a vendor under SDI will be pseudonymous. When the pseudonym management policy of user requires that a new pseudonym be generated for a user, it is necessary to validate the new pseudonym to verify to a vendor that the user has a unique pseudonym for its site.

A user can interact under different pseudonyms with each server, as dictated by her privacy and profiling policies, and the declared policies of vendors with which she interacts. Each pseudonym is associated with a signed PID, (pseudonym ID), a private key that is useful to validate that the PID has not been stolen. The client-proxy also maintains a shared key with each vendor, because symmetric encryption/decryption is cheaper than asymmetric public keylprivate key encryption.

There does not exist a central database of public and private keys for users, or user pseudonyms. We use the method of blinded signatures to certify user-generated key pairs, so that only the client-level proxy servers can link PIDs to User identities. Client-level SDI proxies generate new PIDs, that are unique with a high degree of probability << see UUID reference >> , and blinded before authenticated for use within the system of Secure Data Interchange.

A key feature of the system for administering pseudonyms is that the pseudonym administering authorities cannot build dossiers of the pseudonyms that are authorized for each user, because users submit "blinded" PIDS to be validated. The only information that a PAS has is for each unique user ID, what vendors has the user registered with. This information cannot be linked to the real (physical world) identity of the user because the central SDI user ID server has no information about the user IDs that are authorized for each user that registers with SDI (because the UUID is blinded before signed by the central SDI server).

To generate a new pseudonym for a particular vendor the proxy requests a new PID from the "Pseudonym Administering Server" that has a trusted-agent relationship with the vendor. The proxy provides the PAS with the signed UUID, and if the PAS can verify that the UUID has not already applied for a PID for use with this vendor then the PAS will sign a blinded PID provided by the proxy, with a signature that indicates the vendor that it is valid for— and make a record that a PID has been authorized for user with UUID and vendor V.

Each Pseudonym administering server has a public key /private key pair (PKPAS, SKPAS) for each Vendor for which it validates new pseudonyms. A PAS will sign the public key of a pseudonym using the PAS private key associated with the particular vendor. In effect, the operational pseudonym that a user uses is the triple (S(PK, SPAS), SK, IP), representing a pseudonym that is authorized for a particular vendor server.

There is no information compromise here, other than it becomes possible to construct the set of vendors that a user with UUID has applied for PIDs with. However, the only entity that knows the true identity of the user with UUID is the client-level proxy, because the UUID was blinded before signed by the central SDI server when a new user is registered.

In this way it is possible to provide vendors with guarantees that PIDs are once-in-a-lifetime for a user, so that vendors can continue to achieve at least the same level of personalization as they do without SDI in an SDI system. In this way SDI can guarantee once-in-a-lifetime pseudonyms to those vendors that require persistent interactions with users.

The PID and associated private key can be used to validate an initial information exchange with a vendor, but it is also possible to perform follow up message exchange using a shared key pau—this is more efficient to implement that an asynchronous key pair cryptographic solution. Messages can be encrypted with the shared key, that only the user and the vendor know. This (1) validates that the message is from the sender; (2) ensures that only the intended recipient can read the message. << refer to new Bell Labs paper here >>

3.61 Implementation Details

Figure 11 shows a flow-chart for creating a new "relationship" with a vendor. The Pseudonym Administering servers (PAS) provide for validation of new pseudonyms. Each vendor selects a PAS that will be responsible for managing pseudonyms for its domain. When a client-level SDI proxy requires a new pseudonym for a user UUID, the proxy does the following steps:

1. Request the URL of the PAS from the server for the new vendor that the user wants to initiate a persistent relationship with.
2. Send a message to the PAS with the tag "New Pseudonym", the URL of the vendor, and its validated UUID. Also be prepared to answer a challenge/response with the private key associated with UUID.

The PAS for the vendor then checks that it has authorization to administer pseudonyms for the vendor. The PAS performs the following steps:

1. Check that the vendor URL corresponds to a vendor for which it has pseudonym-management authority.
2. Verify that the W I D is correctly validated for use within SDI (indicating that the user is a member of SDI).
3. Verify that the client proxy server represents the user, through a challenge-response sequence that requires that the proxy has the private key for UUID.
4. Look up the user ID in the database of user IDs that the PAS maintains to check that a pseudonym has not been authorized for this user and this vendor.
5. Send a message to the user, indicating either (OK, or DENY).

Given that the client-level SDI proxy receives an OK response, the client-level SDI proxy then generates a new PID, and private key, and blinds the PID for validation:

3. Generate a new key pair, (PID, SK).
4. Blind the public key, and send a message to the PAS with the tag "Request Certification", the URL of the vendor, the signed unique user ID, and the blinded public key, B(PID)
5. Receives a signed copy of the blinded public key in return, and removes the blinding factor to obtain S(PID, SKPAS).
6. Generate a new pseudonym, from the components, (S(PID, SKPAS), SK, EMAIL), where EMAIL is a new e-mail address for the pseudonym.

The public key that represents the new pseudonym is signed with the private key of the PAS that relates to the particular vendor. This enables a vendor that receives the pseudonym to validate that the pseudonym is unique for the user, to enable persistent interactions across multiple sessions.

3.7 Personalization of information/Manage e-mail

The client-level proxy can provide vendors with "certificates"to enable them to send e-mail to the user (under a particular pseudonym). Outgoing e-mail includes a certificate that a vendor can use to reply, and a pseudonym ID in place of the standard "from" field. The certificate is of the form S(M, SKP), where P is the pseudonym of the user, and M = (PKP, PK*V). The vendor includes the signed certificate when replying to the user.

3.8 Automatic release of personal information

Maintains certificates as well.

3.9 Enhancing profile information

Receives update requests from SDI server. allowing users to access any profile information that is stored in the SDI server, and request changes to that infonnation if it is not accurate, or if the user does not want content to be available to third parties. This information is only released to users via the proxy server that is authorized to represent the user for the relevant pseudonyms.

3.10 Manage iAmWorthIt module/Negotiation

The proxy server determines what kind of site in being visited, from SDI credentials that are provided to a vendor's SDI proxy server, and either embedded in web documents, or provided in prenegotiation. The proxy then determines an appropriate type of interaction level with the vendor, depending on the profile and credentials of the vendor, the type of information required by the vendor, and the user's privacy policy.

The system of Secure Data Interchange allows user's to receive compensation, what we will call "community dollars", in return for providing information to vendors. The architecture supports a negotiation between user's proxy servers and vendor's proxy servers, to strike a deal about information use and compensation. In some cases the exact nature of an offer may not be anticipated, and the user can be contacted directly.

The vendor's host level proxy server is then granted the permission by the client level proxy to receive certain appropriate information. The vendor can use profile information about a user (for a particular pseudonym) to present appropriate services, products, and prices, including custom priced items and promotional offers (as suggested in the co-pending application entitled "System for Customized Prices and Promotions").

3.11 Remote Retrieval of Profiles

The client-side SDI proxy is designed to be configurable from a remote database. This enables a user to maintain a persistent SDI profile across different client machines, for example at work and at home. The profile, pseudonym and key information that represents a user in its interaction with SDI-enabled vendors and information providers during a session can be saved, and then encrypted and stored in a remote database for user name and password access. Alternative technologies include smart card techniques, where the data is stored on a portable device in encrypted form.

3.12 Use-of-data Policy (Access Control Policies)

The Secure Data Interchange system also allows a user to place constraints on how the vendor can interact with the user, and also on what uses can be made of the information that the vendor releases to the central SDI server and other vendors in the system. The user can specify whether or not the vendor can send electronic and/or physical solicitations to the user, and the user can also specify whether data that is released can be used for (any of): solicitation by other vendors, personalization of service should the user visit a vendor's site. For example, a user might require that any information that is released by a vendor to the central SDI server, and then possibly exchanged with other vendors, is only used to personalize the service and products that a vendor offers to the user should the user visit the vendors site, and is not used for electronic or physical solicitations. The use-of-data policy is augmented to the pseudonym-management policy, and defined over classes of vendors.

SDI enables a user to allow limited and secure data sharing between vendor types of its choice. i.e. if I use the same pseudonym for multiple vendors, I am still not saying-OK you can all share my data (although I might choose to say that). I can say-OK, you can use my data to personalize the service that you offer me without knowing what I did with other vendors. Also, vendors themselves can say-here is the data, allow other vendors who are not competitors to access the data, and refine their models, but not actually access details.

<< page 110 -- 113 1995 patent the user can issue strict guidelines to SDI, vendors, and his/her proxy server, about how information that is stored can be used. >>

3.13 Dynamic Privacy Management

When a user clicks to a new URL, then the first-level proxy server first checks its local cache, to determine whether it already has the vendor's classification certificates. The proxy server requests the classification certificates and public key certificate from the vendor if the vendor is not already in its cache. The proxy server continues by verifying the integrity of the certificates, and checking that the vendor matches the ID enclosed in the public key certificate.

The proxy server continues by looking up the pseudonym-management policy for the vendor, and checking whether an appropriate pseudonym already exists in its local database. If the vendor belongs to a new class of vendors that do not yet have a pseudonym, or the vendor belongs to a class of vendors that require one pseudonym each, or the vendor belongs to a class of vendors that require anonymous interaction, then the first-level proxy server continues by generating a new pseudonym.

When a user interacts with a vendor under a persistent pseudonym then the user must have the pseudonym certified by a Pseudonym Administering server, to certify that this is the only pseudonym that the user has registered for the vendor. Each vendor selects a Pseudonym Administering server that will certify all of its pseudonyms. If the user already has a pseudonym for the class of vendors that the vendor belongs to, but the pseudonym has not been certified for this vendor, then the first-level proxy server will have the pseudonym certified. Similarly, if the user requires a new pseudonym for this vendor, then the first-level proxy server will generate a new key pair and have the public key certified by the Pseudonym Administering server.

The first-level proxy server performs the following steps (we use the technique of blinded signatures so that the Pseudonym Administering server does not know the public key that it certifies):

1. Generate a new key pair, (PKP, SKP) - or - lookup an existing key pair that will form a pseudonym for this vendor.
2. Blind the public key, by multiplication with a random number, and form the message M=(S(PK*U, SKSDI), S(B(PKP), SK*U), S(PK*V, SKSDI)).
3. Encrypt the message with the public key of the Pseudonym Administering server, and send the message.

The Pseudonym administering server checks that the unique public key of the user is signed correctly by the central SDI server, and that the user has the private key associated with that public key by using the public key to verify that the blinded pseudonym public key is correctly signed. Finally, the Pseudonym Administering server checks that the public key of the vendor is signed correctly by the central SDI server, and then checks that the user does not already have a pseudonym for the vendor, in a database that it maintains of which unique user PKs have requested pseudonyms for each vendor.

If everything is OK, then the Pseudonym administering server signs the blinded public key for the user's new pseudonym using a key pair that it maintains for the vendor, and returns the signed blinded key, S(B(PKP), SKPAS,V) to the first-level proxy server.

The final step in the protocol is for the first-level proxy to remove the blinding factor, so that the user now has a new key pair, (PKP, SKP), with the public key certified by the appropriate Pseudonym administering server private key, S(PKP, SKPAS,V), to demonstrate that this is the only pseudonym that the user has for the vendor.

The first level proxy server now connects to the vendor in secure-SDI mode, by sending the vendor the certified public key for the user under this pseudonym that will identify the user to the vendor. The proxy server continues by sending the data privacy policy for the user with this vendor, signed with the private key for the pseudonym. This serves three main purposes: (1) it demonstrates to the vendor that the user is indeed the owner of the public key PKP, because the user has the secret key that is associated with it; (2) it informs the vendor about the data-privacy policies that the user requires; (3) whenever the vendor submits information about transactions with this user to the central SDI server it must also submit this certificate to verify that it is following the user's guidelines.

3.14 Dynamic Profile Management

When a user connects to a site and provides a certified public key, the first-level proxy server also provides a time-stamped certificate of connection, S((PK*V, T), SKP), where T is the current time, PK*V is the unique public key of the vendor, and SKP is the secret key of the user for the pseudonym that it uses with the vendor. This "connection certificate" is used by the vendor to request a profile-release from the central SDI server.

In addition, the first-level proxy server also provides a basic-profile to the vendor. The basic profile for a user contains no identifying information, but can contain whatever general information the user is happy to release across all pseudonyms, such as the user's age, nationality, state, sex. This is the basic profile that is configured by the user during initial registration with SDI.

The vendor provides this certificate to the central SDI server, and requests profile information about the user with pseudonym P that is stored within the central SDI server, and is authorized by the user to be released. If the vendor is authorized to receive dynamic profiling information, such as the recent web footprints of the user, and the material that the user has been reading, and the physical location of the user, then this information is released to a vendor when the user connects to a new site, according to the dynamic profile policy of the user.

3.15 Merging off-line and on-line data

Figure 12 shows a flow-chart for how the central SDI server can request that new information be merged with a user's on-line profiles.

The central SDI server can associate off-line information about a user with a user's on line pseudonymous profile, even though the central server does not know the user's pseudonym IDs. This can only be done with the user's consent, and may also involve appropriate compensation. Within the system of iamworthit (see section XX) we can credit users for both off-line and on-line information.

Merging a marketing database with SDI usr-profiles can be useful both to initialize the database, for example when asking a user questions to generate an accurate user profile rapidly and efficiently. Off-line data can also add useful richness to on-line profiling information, which may be largely contextual and low on details/factual information. For example, off-line data can include information such as whether a user owns a car, rents an apartment, has house insurance, life insurance etc.

SDI can also extrapolate correlations to other user profiles, on the basis of common SDI-profiles, for example using statistical techniques.

It is often the case that individual customers appear in some databases, but not in others. Under normal circumstances, an analyst working across different databases would be faced with a large number of incomplete customer records, each with gaps corresponding to the fields of the databases to which they don't belong. A solution to this problem is offered by SDI, which is capable of drawing correlations between different databases - this information can be used to generate predictions to fill in the gaps of incomplete customer records. The result is a full set of customer records that can be meaningfully sorted or filtered by any of the combined fields, and which can now be handled as a unified set of data, suitable for use by standard database analysis systems.

In a typical example, SDI might be used to combine a demographic database, such as the one offered by the Econometrics Corporation, with a commercial database, such as the one offered by Claritas. The Econometrics database consists of 180 million different customer records, but at a fairly coarse-grained level of detail, consisting of such information as age, gender, family status, location (at the state, city, or zip code level), and personal income. In comparison, Claritas offers a smaller base of customers, but includes information of arguably higher quality, since it breaks customers down to the geocode (sub-neighborhood) level, and includes much more detailed information on personal spending habits across hundreds of different purchase categories. A logical reason to combine these databases would be to supplement information about customers in the vastly broader demographics dataset with particular predictions about their personal preferences and likely commercial spending habits. One could imagine using this augmented data set to support a web site that instantly customizes itself to new visitors' preferences. Since the number of records in the Econometrics database is equivalent to roughly 72% of the population of the United States, it is likely

that most first-time visitors to the site will already have a "thumbnail sketch" in the system, and can thus be greeted with an page appropriately configured to their personal tastes.

The technical details of the combination process (which have been described elsewhere in the patent) to a large degree depend on the amount of overlap between the databases, that is, the number of customer records which are shared in common.

Suppose the demographic databases' fields are coded (x1, . , xn), and the commercial databases' fields are coded (y1, ., yn). Suppose further that customers in set A appear only in the demographic database, customers in set B appear only in the commercial database, and customers in set C appear in both.

The process of supplementing the fields of customers A depends completely on the derivation of the distribution f(y1, ., yn | x1, . , xn), which describes the correlation of fields in the commercial database on fields in the demographic database. As previously discussed in the patent, different techniques may be used to create this distribution, depending on the size and variety of C.

As a concrete example, one could imagine that set C includes customers from rural areas. The demographic database would reveal that, although their incomes aren't huge relative to the national average, they tend to spend a lot of it (i.e. are active consumers), have large families, and purchase large vehicles. The commercial database might show that they enjoying hunting magazines and Ford trucks. If they live inland, they buy hunting equipment, if they live near the ocean, fishing equipment.

If these trends are dominant in set C, they will impact the distribution function. Thus, when a browser from a small town in Texas with a typical income pattern visits the automated website, he could be greeted with discounts on truck accessories and a small sidebar with news on the hunting season. On the other hand, a visitor from a small town in Maine might be given the same truck discounts, but would have news on the fishing season.

Although the demographic dataset is arguably the weaker of the two in terms of content, the fact that it contains even a small amount of information on most people in America makes it very valuable for handling first-time visitors, since most of them will appear in it. By using SDI to leverage the more detailed information in the commercial database, we are able to supplement the rough demographic data with predicted commercial preferences. This allows us to construct more detailed thumbnail sketches for each customer, allowing our reception of first-time visitors to be much more appropriate (since knowing personal hobbies or interests tells us much more about a person than general income level).

4. *The ISP-level Proxy Server*

The ISP-level proxy server is positioned just behind the firewall of the user's local dial-up network (ISP or Intranet). The proxy provides protection for users operating under pseudonyms from point-to-point attacks and HTTP header-tracking by stripping HTTP header-information and forwarding HTTP packets on to their destination with no information other than their source at the ISP-level proxy server. The ISP-level proxy also supports pseudonymous e-mail, between users, and between users and vendors.

Figure 2 shows a couple of users connected to clients, that are in turn connected to the Internet through a local intranet, such as the network of an Internet Service Provider (ISP).

The proxy "washes" outgoing messages of any information that would compromise a user's pseudonymity, for example the "referral" field that contains the previous URL of a user in a HTTP message. HTTP messages also leak other information, for example browser software on a user's client machine, the operating system and a user's IP address.

4.1 Support for Pseudonymous electronic mail

A user can receive electronic mail through the PID and associated IP address of the ISP-level proxy server.

The preferred implementation of this system allows the user to periodically check for new mail. The client-level proxy gains access to the mail box that is associated with a pseudonym by providing a correct response (signature) to an ISP-generated challenge. Notice that with this solution, the ISP-level proxy has no way to connect the pseudonyms of a

user, so long as the user's client is not identified in its messages to the ISP-level proxy server other than by the PID that the proxy makes a request for.

** DP. Must be careful to "wash" HTTP at the client-level proxy as well % as the ISP-level proxy.

We can extend this mechanism using a technique taught in the Lucent Personalized Web Assistant. The LPWA provides for a sequential access mechanism to the mailboxes that belong to a user through a one-way function that takes the user's SDI log-in name and password, and an integer from 1 to N, and computes the mailbox location. The mail server does not need to maintain a list of pseudonyms for each user, because the user is able to efficiently access all of its mailboxes sequentially as a function of other information.

Another variation, that relies on the user placing trust in the ISP-level proxy server, provides the ISP-level proxy with the e-mail address for each pseudonym. This push method is more efficient, because the ISP proxy and the client proxy communicate only when new messages arrive, but provides the ISP proxy with information to compute all the pseudonyms for a single user—probably undesirable.

4.2 Support for anonymous profile-based search

The anonymous profile-based search allows a user to release her profile as an addition to a query term to an general search engine (such as Altavista or Yahoo), and have SDI perform additional filtering of the results of the search to refine the pages returned on the basis of their profiles and the user's profile. This is an example of how SDI leverages existing Internet technologies. Figure 9 shows how the anonymous profile based search can be implemented on the ISP-level SDI proxy.

4.3 Maintain User Profiles

4.4 Support Electronic Payment/Physical mailing solutions

5. Vendor-level SDI Proxy

The vendors are represented with servers that provide an on-line electronic commerce store-front for the services or products that a vendor provides to users. The servers may connect through their own pseudonymizing proxy servers, but more likely a vendor will not require pseudonymous interactions with users, and will be happy for users to build dossiers on interactions with the same vendor. The vendors make use of the support for pseudonymous electronic commerce functionality that the Secure Data Interchange system supports - for example, pseudonymous payment mechanisms and physical mailing of products and letters to pseudonymous users. The vendors subscribe to the services of SDI, and can receive payments for user information that they provide to the interchange.

Vendors connect to users and other vendors through SDI proxy servers. The vendor-level SDI proxy can profile users, periodically release information to the central SDI database, and support various statistical generation modules for analysis of profiles collected and requested from other sources. Figure 1 shows a typical vendor that is connected to SDI through a server and vendor-level proxy server.

5.1 Profile Building

The main function of the vendor-level SDI proxy is to build a deep profile of users as they interact with a vendor

5.2 Profile Release

and manage the release of information to the central SDI data warehouse. Figures 3, 4, 5 and 6 show patterns of data flow to and from the vendor.

5.3 Request Profile Updates/ Cache metatags/ Manage target object profiles.

5.4 Markup Web documents with Profile information and Personalization commands

A new vendor may choose to operate under different pseudonyms for different users or classes of users, but in general we will assume that a vendor has a unique key pair, and that it is happy to assume the public key of its key pair as an identity, in addition to its more traditional corporate identity. In what follows we assume that the vendor will operate under a single public key with all users, although the extension to multiple pseudonyms is a trivial extension, just using the same technology as described for users.

When a new vendor wishes to register with SDI, the vendor configures a first-level proxy server to manage its SDI functionality. The first-level proxy server is similar to the proxy server configured by new users. Initially the proxy server will generate a new key pair for the vendor, (PKV, SKV), and submit the public key to the central SDI server for signing, along with an identifying code, such as the IP address from which the vendor operates, ID. The vendor receives S((PKV,ID), SKSDI) in return. In addition, the vendor is audited by SDI and is certified according to the class of products or services that it provides, receiving message S(M, SKSDI), where M=(L(V), PKV). The vendor is also certified according to the anonymity of the service or goods that it provides, receiving either an anonymous certificate (e.g. if it deals in high-volume, non-personalized goods), S(M, SKSDI) where M=(PKV, ANON), or receiving an non-anonymous certificate (e.g. if it deals in low-volume, personalized goods), S(M, SKSDI) where M=(PKV, NON_ANON). The vendor will provide these three certificates whenever a user connects to the page and the user does not already have the certificates cached for the vendor.

A vendor's data privacy policy can also usefully include the placement of restrictions upon any portions or all of their data disclosed to the interchange, such as which other vendors (or vendor types) are entitled to receive the advertising benefits and what types of advertising benefits derived from hislher data are they entitled to; for what advertising purpose can an "entitled" vendor use it for (if it is not deemed "confidential" the disclosing vendor may request beforehand a copy of the solicitation to hislher customers); limiting use of its data exclusively to other vendors who through their data contributions can provide significant reciprocal benefit to them (often occurring in dealing with competitors); physically confining the data to the interchange (without transferring it to the other vendor); data may only be transferred to non-competitors; advertising solicitations that result from the data is of a form that is non-competitive; recipients of data must remain non-competitors.

For example, a vendor can request that the commercial domain of the requesting vendor not match certain vendor types. Vendors must register their business type with the data interchange, and violators can be punished. Alternatively, data can be constrained only for vendors that are improving existing data models, but not to vendors that are seeking new customer prospects. Disclosing vendors may also require that their advertising campaign and possibly even their corporate identities remain secret.. Because SDI is a comprehensive repository of statistical data (and customer rating information across a variety of criteria) about sites, it is able to deploy data mining techniques, provide highly detailed statistical information and verification as to the nature of each vendor's business and its customers. Thus helping SDI to better identify, reclassify and enforce this data privacy policies. Provided that the vendor has agreed to release such information to another type specific vendor, this information can serve as additional useful information (particularly quality and other rating credentials about vendors) to help vendors do a better job of selecting vendors by which they establish their data privacy policies.

6. Central SDI Server

Also shown in Figure 1 is the central SDI server, "SDI data warehouse server". The central SDI server maintains extended profile information for each user, at a pseudonym level, and profile information for each vendor. The central SDI server also supports cross-vendor and single-vendor personalization tools, such as multi-attribute collaborative filtering techniques. The data warehouse operates on a trusted and secure server. The server is able to provide information to vendors that enables models to be enhanced, and synergies between data sets realized, often without providing vendors with direct access to the actual user profiles (unless the user and submitting vendor has authorized such a disclosure.)

The central SDI-server maintains profile information for each pseudonym of every user registered with SDI, aggregated from information provided by multiple sources (e.g. vendors and clients). The server contains modules to perform data mining and collaborative filtering on profiles, and release results.

The SDI server ensures the integrity of data, and prevents data being used for unauthorized purposes. The server performs collaborative filtering analysis on the large data sets, and releases vendor-specific model enhancements. The central SDI server will also provide targeted mailing to a virtual mailing list of targeted users, without releasing the pseudonyms to a vendor. The main SDI server also maintains an encrypted portfolio of pseudonyms and personal information for each registered user, and is configured to release that information to client-level proxy servers when a user logs-on to SDI. Another important roles of the SDI server is the categorization of vendors' business interests and data-handling policies.

6.1 Profile Updates

Figure 3 shows modes of data flow to the central SDI server.

SDI will accept updates to profile for pseudonym P by communication from the client-level SDI proxy server associated with pseudonym P. A user can log new information about one of its pseudonyms by sending a request R, with profile updates, to the central Secure Data Interchange server. The user's identity is verified using public-key cryptography, managed at the client-level proxy. A user can also reveal the equivalence of a set of pseudonyms to SDI.

The Secure Data Interchange server will accept updates to the profile of a pseudonym from any party in the system that has registered with SDI. The user can request to see the profile for any of its pseudonyms at any time, to check that the information accurately reflects his/her behavior, interactions and transactions with vendors and other parties. A user request for a profile is initiated by a message M from the user. The message M contains a one-time key D, with which SDI will encrypt the profile. Only the user knows the key. The message M is encrypted with the public key of SDI, and then sent to SDI via the user's trusted proxy server.

SDI verifies that the request for profile disclosure has been received from the authorized proxy server for pseudonym P, and then sends the profile, encrypted with the one-time key D, to the user via the proxy server.

SDI records a vendor ID when a party other than the user makes an update to the profile of a user's pseudonym. The vendor must be registered with SDI, and make a signed commitment that it will only register information that is accurate and collected in good faith. Whenever a user challenges information, SDI can trace the vendor that provided the information.

6.2 Maintaining Data-use Guidelines

6.3 Use of Data

Vendors can check in data to SDI in any of the following forms:

i. Data or Randomized Data, with/without pseudonym
ii. Aggregate information on profiles in user-base, with/without pseudonyms

Suppose that a user has requested that a vendor does not reveal any information about the transactions that a vendor has executed with the user (under a pseudonym). The vendor can still reveal randomized information that will in no way identify a user (even though the user interacts under a pseudonym). We can perform system-wide collaborative filtering on randomized information (details below).

SDI can reveal information to vendors in a number of forms:

i. Data or randomized data, with /without pseudonyms
ii. Analysis results: For example

if a customer is interested in product X, suggest product Y
if a customer with profile type A hits your site, suggest

product Y
if a customer with profile type A is interested in product
X, suggest product Y
if customer P hits your site, suggest product Y
if customer P hits your site, and is interested in product
X, suggest product Y
The SDI server can provide analysis based on the profile of a user with pseudonym P without requiring that the vendor can access pseudonym P by providing a secure function evaluation procedure, that analyses the encrypted profile and provides results (details below).

The SDI server does not actually reveal any private information about the user to the vendor, so long as the results of the analysis are restricted to recommending products that the vendor has that might be just what the user is looking for. We protect the privacy of users while providing personalization. The options above provide for a whole range of personalization, from very coarse-grained, to very fine-grained, but it is all dynamic and secure. Furthermore, we can provide the ability of a vendor to personalize service to users that hit his/her site, but prevent the vendor from soliciting users.

SDI can also retain control over information on users by placing cryptographically enforced "expiration dates" on the ability of vendors to analyze profiles that are delivered with requests from users. This allows sophisticated pricing models (see later section on pricing models).

There is a hierarchy of personalization that is available to each vendor in the system. As a base level, a vendor can request that SDI performs data analysis on its current user base, and with only transaction information collected at its own site. The next level performs analysis on the vendor's user base, but also drawing on profile information from across multiple vendors' sites. The next level allows vendors to personalize service to incoming users on the basis of an analysis of the profiles of the new users. Finally, SDI may provide a list of pseudonyms and recommendations, so that the vendor can target new users.

6.4 Aggregating Profile Information from Clients and Vendors

6.5 Releasing Anonymous Profile Information for SDI-enhancedSearch

Figure 9 illustrates how anonymous profile-based search can be performed on ISP-level SDI proxies, with information provided from the central SDI database.

6.6 Data Mining/Collaborative Filtering Functionality

One of the core purposes of SDI is to provide a common location and format for information that has been gathered from a wide variety of sources and that might require different sorts of analysis. Since its framework is designed to generically handle different types of data and algorithms, SDI can be used as a platform to explore and exploit the rich connections that potentially exist within and across the databases of different vendors and customers.

This section describes in greater detail the types of data, analytical methods, and forms of validation that are available to SDI. The Secure Data Interchange architecture in its preferred implementation integrates the architecture issued U.S. Patent No. 5,754,939 "System for Customized Identification of Desirable Objects" into a system for secure data exchange between multiple parties. The aforementioned patent teaches a method for profiling objects and users over a bi-directional distributed network, such as: an ISP, multiple ISP networks, a Web hosting network, or server software (such as data mining or recommender software) that is linked to a coalition of sites (such as a portal or Internet mall). The current invention, the system of Secure Data Interchange, allows correlations to be identified between vendor's data sets, that allows accurate profiling through the application of statistical methods, without providing vendors with explicit access to the profiles of users – because profiles are provided in anonymized and randomized forms. There are less efficient methods that can be used to identify correlations, for example using customer demographics, and vendor categories, to suggest which vendors might be well placed to form dynamic syndication relationships. With SDI it is possible to leverage as many data sources as are available, about users and the

target objects with which they interact. In fact the degree of the measure of improvement in predicting user behavior (or increasing clickthrough) is approximately in direct proportion to the square root of the number of user profiles and target profile interest summaries which are known. The emphasis in the aforementioned patent is on the bilateral relationships between vendors and users, and the architecture is not designed to support secure and privacy-protected data interchange and analysis across the user bases of different vendors.

In the system for SDI we push control of the profile for each user to the client software that runs on the machine local to the user, and provide for personalization through dynamic processing of information on the user's client machine. We enable vendors to exchange data sets only to the degree that is mandated by users, and provide technical solutions to enable significant leverage of data while maintaining user privacy.

With SDI we enable second-level proxy servers for vendors to interact, and also make available the user profile information that is collected at the client-level and ISP-level SDI proxies to develop more accurate profile information, through the combination of "deep" vendor-specific information with broader network-level information. The ISP-level SDI proxies can also capture some information (in a privacy-protected form) about users and vendor sites that do not subscribe to the SDI service. The goal of the architecture is to continue to allow relevant filtering to be applied at sites that a user has never even directly visited before, or interacted with before, without that site accessing the profile information of the user.

The supporting architecture as stated in the above referenced patent also allows for profiling statistics to be collected and processed in such a manner that the information and vendor servers may both contain and implement the modules for profiling in a distributed manner. In the present invention the profile generation capabilities are implemented at various levels – in particular, at the second-level proxy servers of ISP networks and vendor servers, and within the central SDI data warehouse.

Web page tags include profiles of target objects, user quality ratings based upon overall quality as well as other criteria (e.g., value, price, entertaining, informative graphic/visual appeal, etc.) and data mining analysis of each criteria as it corresponds to target objects, target object attributes, user attributes (and vice versa), trend analysis statistics, location data (for target objects representing physical or geographical items). User information, in addition to profiles, can include data mining and trend analysis statistics, user provided ratings for target objects, and resolution credentials.

Vendors can use queries, data mining, menuing, and other techniques (such as described in the aforementioned patent), in order to gain access to desired user profiles that the vendor is allowed access to. The information that is stored against a user's pseudonym in the central SDI server is checked against the data-release certificate that the user provided to releasing vendors, and is therefore data that the user has indicated can be stored on hislher behalf.

6.61 Structure of the Central SDI database

The central SDI server is structured as a relational database, with data that is submitted by vendors and users to the server indexed by pseudonym. The basic structure of a data record relates a pseudonym ID to a vector of numbers, representing the profile for that pseudonym. (Pseudonym–id, Profile). We also add tags to indicate the usage-restrictions on the profile information, and whether or not the data is randomized.

There are several types of information which can characterize both users and items. SDI is intended to function as the intermediary between a vast web of vendors, on the one hand, and individual consumers, on the other. The major sources of data used by this system are therefore:

1) Demographic. Such data will most likely be elicited by SDI from vendors and consumers when they initially register for the service, and details very general characteristics about them. It will consist of numbers and categorical values (age, zip code, sex, level of education, etc.).

2) Commercial. This is the kind of data any that vendor collects in the course of doing business (especially e-commerce); generally, it links customer codes to purchase items, dates, quantities, and prices. Depending on the nature of the business, this data could be fairly complex, and might well include text. For example, one could imagine that a bookstore, in addition to keeping track of its sales history, collects book reviews, author profiles, and plot summaries.

3) Behavioral (vis-a-vis the Internet). When a customer accesses the Internet through an Internet Service Provider (ISP), his ISP is in a unique position to monitor and record every behavior he exhibits while browsing the World Wide Web, writing email, and uploading/downloading files. This information is fairly complicated, as every site visit involves the creation or consumption of content (with associated sound, images, and text), the generation of queries (when using text-based search engines), and the navigation of hyperlinks. An additional complication is that such data is dynamic, in that differing amounts of time are spent for on-line activities and might well reflect the customer's personal preferences (e.g., a long time spent lingering between two choices in an on-line catalog might indicate an equal level of preference, whereas a speedy exit from a site might indicate general lack of interest).

Further information is provided by XML tags attached to pages browsed by the customer. The mere presence of such tags allows for correlations to be drawn between different web pages (e.g., a common XML tag used by travel-related sites), because it implies similarity. Furthermore, it is conceivable that such tags could encode more refined measures of a web page's content, such as browsers' evaluations of its value. For example, a web page of interest to scale modelers, in addition to having images and text related to model trains, might have an XML tag that shows that other scale modelers have given the web site a "five-star" rating. This page should therefore be given a greater weight when SDI is used to create correlations of interest to model hobbyists.

To fully represent these different sets of information, SDI handles collections of data of the following types:

1) numerical (e.g. an age, price, or period of time)
2) categorical (e.g. a color or musical genre)
3) text

A common task for SDI is to compare and correlate different customers, which might well be represented by mixed collections of numbers, categories, and blocks of text. This is handled by treating each customer c_i as a vector in a space whose coordinates correspond to the fields of data available. In the following description we refer to a customer, but when a user interacts with a vendor under a pseudonym, the profile information will only relate to information provided to the central SDI server for that pseudonym.

If there are m numerical pieces of data available, there will be n corresponding coordinates in the data space, $(x_1, \ldots, x_m)$.

For each category i, there will be a corresponding number of values, n_i. Hence, for a color category {red, white, blue), $n_{color} = 3$. Since each value is assigned its own coordinate, category i is represented as an n_i dimensional vector, y_i. Hence, the total number of dimensions used to describe the full set of n categories $(y_1, \ldots, y_n)$ is

$$\sum_{i=1}^{n} n_i$$

Note that sparse methods are especially useful here, since a categorical vector y_i will typically consist of mostly zeroes, with a single non-zero coordinate representing the categories' value (i.e., we encode the color red, using the previous example, as (1,0,0)). Note also that category vectors with different values are treated as orthogonal by the system.

A final issue is the representation of text. As described in previous related patents, all relevant blocks of text in the database are converted into a dictionary that maps unique strings to the number of times they appear in the database. An appropriate TFIIDF weighting function is chosen and calculated for each of the p words that appear in the dictionary. The full set of text connected to a single customer can thus be represented as the vector $(z_1, \ldots, z_p)$, where each zi equals the number of times the word i appears in text related to the particular customer multiplied by the TFIIDF score assigned to word i.

In summary, when a database describes its customers using a combination of numerical values, categories, and text, customer i can be represented by the vector $c_i = (x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_p)$.

6.62 An Example Profile Vector

Suppose we have a database containing information on customers' ages, their musical preferences (i.e. an answer to a survey asking: "Which do you prefer, Mozart or the Beatles?')), and the contents of the emails they've written. Furthermore, suppose the only salient variables in all the emails written consist of the words "Beatles", "Mozart", and "practice", and that we are using the function

$$TF / IDF(x) = 1/\sqrt{n_x}$$

Where n_x represents the number of times word x appears in the dictionary. So, if the word "Beatles" appears a total of 217 times across the full set of customer emails, $n_{Beatles} = 217$, and

$$TF / IDF(Beatles) = 1/\sqrt{217} = 0.067$$

We now want to represent one of the customers in the database; he's a 10-year-old boy who prefers Mozart to the Beatles, and who wrote an email to his friend that mostly describes his attempts at practicing Mozart, but in passing mentions his sister's new Beatles CD. Suppose he uses the word Mozart 2 times (although it appears 456 times in the full database of all customers' emails), the word Beatles 1 time (appears 217 times in database), and the word practice 3 times (appears 77 times in database).

We define the following coordinates:
x_1 = age = 10
y_1 = {Mozart, Beatles) = (1,0)
z_1 = # of times customer uses word "Beatles" x TF/IDF("Beatles") = 1 * 0.067 = 0.067
z_2 = # of times customer uses word "Mozart" x TF/IDF("Mozart") = 2 * 0.047 = 0.094
z_3 = # of times customer uses word "practice" x TF/IDF("practice") = 3 * 0.114 = 0.342

In our example, then, we might encode this boy as customer 1:
$c_1 = (x_1, y_1, z_1, z_2, z_3) = (10, 1, 0, 0.067, 0.094, 0.342)$

6.63 Choosing an Appropriate Level of Data Granularity

We define the term granularity to denote the level of detail available within a given set of data, which is often structured hierarchically. Suppose a grocery store database contains records for a box of flavored gelatin powder. This could be categorized in a variety of ways; moving from the most specific to the most general, we might treat this data point as "12.5 ounce, strawberry flavor, Jello-brand gelatin dessert" (which would be entirely different from "12.5 ounce, banana flavor, Jello-brand gelatin dessert"), or as "12.5 ounce Jello gelatin" (a categorization which would treat as identical the strawberry and banana Jellos), or as "flavored gelatin", or as "dessert", or as "food", or as "grocery".

When analysis is performed on such data, the level of granularity chosen will have a strong effect on the outcome of the analysis. If the level of granularity is too fine-grained, the data will be too sparse, although it could be potentially aggregated to the next highest level of granularity. If the granularity is too coarse, the results of the analysis might be overly general (e.g., a customer would find a collaborative filter useless if the only recommendation it makes for a dessert choice is "go to the grocery section of the store").

Since the level of granularity will have a salient effect on the outcome of an analysis, it should be chosen very carefully, and might well play a factor in pricing when a vendor chooses to sell its data.

6.64 Methods used for data analysis

In order to perfonn a wide range of analytical tasks, SDI needs to make use of a variety of computational approaches. These are described below, starting with the simplest methods first.

- (1). Standard Database Searches

Since most of the data will be stored in centralized databases, simple searches, queries, and data filters can be implemented by means of standard SQL commands. Typically, data will be collected or sorted using efficient database calls before being fed through analysis routines; once complete, the results can be fed back out to the database environment for further efficient manipulation.

- (2) Metrics – Measuring the Similarity Between Profile Vectors

Given two customer (or vendor) profiles, c_i and c_j, it is frequently desirable to know how similar they are. For this purpose, we define the similarity metric $M(c_i, c_j)$ to be a function that takes as input two customer vectors and returns as output a numerical value in the range [0,1]. When two customers c_i and c_j are identical, $M(c_i, c_j)=1$; when they're completely different, $M(c_i, c_j)=0$.

The problem is somewhat simplified by the fact that we treat all customers as vectors. Given two customer vectors, we can use the correlation between them to serve as our metric:

$$M(A,B) = \cos\theta = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

Note that θ here represents the angle between the vectors A and B, and that we expect all coordinates of the vectors to be positive (in order for M(A,B) to keep its output in the range [0,1]).

In more complicated cases, however, a customer vector might contain multiple fields with varying ranges of values. For example, we might have customer vectors of the form $c_i=(age_i, income_i)$, in which the maximum age is 80, but the maximum income is 300,000. In such cases, the coordinates with larger values will dominate the similarity metric, overwhelming any influence that smaller fields might have.

This requires a normalization of the customer vectors, which can be done in several different ways. One approach would be to scale every coordinate by the maximum observed value, forcing all coordinates to lie between 0 and 1 (again, enforcing the rule that all coordinates must be positive).

$$c_i = \left(\frac{age_i}{\max(age)}, \frac{income_i}{\max(income)} \right)$$

The only problem with this is that if a coordinate's maximum value is an outlier (being vastly bigger than the typical value), most of the coordinates' values will seem unusually small once they are scaled by the maximum. In such cases, it might be better to scale the values with a "squashing" function such as the sigmoid, which deadens the impact of extreme values; one such configuration would be the following:

$$\overline{age_i} = \frac{age_i - mean(age)}{\sigma_{age}}$$

$$\overline{income_i} = \frac{income_i - mean(income)}{\sigma_{income}}$$

$$c_i = \left(\frac{e^{\overline{age_i}}}{1+e^{\overline{age_i}}}, \frac{e^{\overline{income_i}}}{1+e^{\overline{income_i}}} \right)$$

Note that the mean and variance of the data points are used to fully normalize them, such that the sigmoid function will spread the values somewhat more evenly between zero and one.

The previous approaches are especially useful for single numerical fields, which might well overwhelm each other if some sort of normalization isn't performed.

A different problem arises for text or large categorical fields, since they can potentially consist of hundreds of coordinates capable of overwhelming the influence of single numerical fields. Suppose we believe the age of a customer is as important as the text of articles read. In such a situation, the thousands of coordinates devoted to the text field would dominate the metric's behavior, negating any influence that age would have on our measure of similarity – clearly not a good situation.

A solution to this would be to find the correlations among the fields taken separately, then average the result. That is, if each customer $c_i = (age_i, text_i)$, where $text_i$ is a vector with a very high number of dimensions, we could define the metric:

$$M(c_i, c_j) = \left(\frac{corr(age_i, age_j) + corr(text_i, text_j)}{2} \right)$$

Where

$$corr(c_i, c_j) = \frac{c_i \cdot c_j}{\|c_i\| \cdot \|c_j\|}$$

The result is a metric that gives equal influence to each field.

- (3) Forming Vectors Into Groups

The process of classification is essential to collaborative filtering, as it allows different vectors to be formed into groups based on some measure of similarity. If we are able to create groups of customer vectors, for example, we can then give individual customers recommendations based on the patterns of their group-mates, who presumably have similar tastes.

K-means Clustering and Nearest Neighbor algorithms are extremely useful for grouping purposes: previous iReactor patents <<*FRED* – references?>> give a full and detailed description of our customized versions. This section gives a brief overview of these methods.

(3.1) Clustering

K-means Clustering is an algorithm used to partition a coordinate space such that all vectors in a given partition are more similar to that partition's vector average (the centroid), than to the centroids of any other partition. It is a process that iterates over the following steps:

0. "Seed" the coordinate space with the initial centroids, which are vectors used to describe the centers of the clusters, in the sense that they are the average of all the vectors currently assigned to the partition. This can be done randomly (assigning centroids random coordinates) if no other information is available, or it can be guided by pre-existing information. For example, if we wish to cluster vectors of music customers, we can use information about musical genres to create initial partitions that correspond to pop, gospel, classical, etc. This will locate the centroids in well-spaced intervals across the coordinate space.

1. Assign vectors to the most similar centroids. This is done for each vector by scanning across all centroids and calculating similarity M(vector, centroid$_i$); once finished, the vector is assigned to the cluster whose centroid has the greatest similarity. In this stage, vectors may switch their allegiance from one centroid to another, if the relative distances to the vector have changed sufficiently since the previous iteration. If no vectors change their allegiance, the iteration process is complete, and the algorithm stops.

2. If the iteration is not complete, recalculate the centroids by setting them equal to the average of those vectors that have been assigned to them. Go back to step 1.

Once the algorithm converges, the vectors are grouped into clusters. The centroids' coordinates as well as the identity of cluster members is useful information that can be passed on to subsequent stages of analysis.

(3.2) Nearest Neighbor

The nearest neighbor algorithm, simply stated, creates a list of those vectors in a database that most resemble a particular target vector. This is accomplished by comparing the target vector, in turn, to every other vector in the database; the similarity between them is recorded, and once the comparison loop is complete the list of similarities is sorted. The top k members of this list are returned as representing those k vectors which most resemble the target.

- (4) Generalizing Across Databases

One of the most useful aspects of SDI is that it allows for inferences to be drawn across different databases through underlying connections in membership or content. An especially strong link can be made between commercial databases if they have customers in common. However, for reasons of privacy, individual customers may choose to use different pseudonyms when dealing with different vendors. This might be preferred by the individuals, but it weakens the inferences that can be made between fields occurring in different databases.

The techniques chosen to infer correlations across different databases will depend on how many pseudonyms are shared in common. At one end of the spectrum, every customer uses a single pseudonym for all transactions,

and makes an appearance in every database. At the opposite end of the spectrum, every customer uses a different pseudonym with every vendor, and may appear in only a single database.

Case 1: All customers use a single pseudonym, and appear in all databases considered.

This is the simplest situation to handle. Since all customers appear in all the databases, the customer vectors' fields are essentially scattered across several locations, but can be easily reconstructed. For each customer, we define a new data vector that concatenates that customer's representation from across the different databases.

Hence, if we are considering databases A, B, ..., Z, and customer i appears in each one, we define a new vector $c_i = (c_{Ai}, c_{Bi}, \ldots, c_{Zi})$, where c_{Ai} is customer i's vector in database A. We then proceed as usual, making inferences with these augmented customer vectors.

Case 2: Most customers use a unique pseudonym, and frequently appear in different databases.

In this situation, although we see some connections between the databases, many pseudonyms appear in only a single location. Using Bayesian techniques, however, we can still make predictions for customer vectors across databases.

Suppose we have a set of databases, A, B, ..., Z. Taking each database in turn, we cluster it using all available data. Thus, using every record in database A, we group A's customers into clusters

$A_1, A_2, \ldots, A_n$. Taking database B, we create clusters using all of B's information, creating customer clusters B_1, $B_2, \ldots, B_m$, and so forth.

Now, scan both databases for common pseudonyms (representing those customers who have interacted with both vendors under the same pseudonym) and create count variables w_{ij} to represent the number of pseudonyms that appear jointly in A_i and B_j.

$$P\left(B_j \middle| A_i\right) = \frac{P\left(B_j \wedge A_i\right)}{P\left(A_i\right)} = \frac{w_{ij} \, I \, total}{\sum_{j=1}^{m} w_{ij} \, I \, total}$$

$$total = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}$$

We can now produce the probability that a pseudonym appearing in A_i will appear in B_j:

For example, if we have a database of airline ticket purchases and a database of restaurant visits, we can create clusters, in the first case, of customers who travel to similar destinations, and in the second case, of customers who eat at similar restaurants. Given that a particular customer belongs to a cluster of people who frequent Caribbean restaurants, we can infer which travel packages would most appeal to him based on the linking probabilities, as defined above.

- Multivariate Extensions:

If we have a third database C, and there are a large number of pseudonyms common to A, B, C, the above probabilities can easily be extended. For example, knowing that a customer appears in A_i and B_j, we can calculate the linking probabilities to any C_k:

$$P\left(C_k \middle| A_i \wedge B_j\right) = \frac{P\left(A_i \wedge B_j \wedge C_k\right)}{P\left(A_i \wedge B_j\right)} = \frac{w_{ijk} / total}{\sum_{k=1}^{p} w_{ijk} / total}$$

$$total = \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}$$

Or, if there aren't many pseudonyms that span all three databases, the probability of C_k given that a pseudonym exists in A_i and B_j could be approximated by:

$$P\left(C_k \middle| A_i \wedge B_j\right) = P\left(C_k \middle| A_i\right) \cdot P\left(C_k \middle| B_j\right)$$

Case 3: All customers use several pseudonyms, and none appear in different databases

In this situation, there are no common customer codes that can be used to create links across the databases. However, the mere fact that several databases have been brought together for analysis should imply that there are semantic commonalities in the data.

Although each database contains different fields, it may be the case that those fields deal with related subjects. A human expert, knowledgeable in the content of the databases, the subtleties of the domain, and the overall goal of the analysis (e.g. the creation of recommendations), will be in a position to create a "common-information profile" that spans the databases. In essence, the common-information profile defines a format that allows vectors from different databases to share a common coordinate space.

The idea is this: the expert designs a high-level vector format that embodies the content deemed important for the project goals. Next, for each database he develops a mapping that encodes the database's elements into the generic format. Finally, the desired analysis is performed on the full set of common-information profiles.

Although the expert will have to create completely new fields for the common-information profile, certain types of data will map directly to the common-information format. In particular, if every database contains text (catalogued and counted, for TF/IDF purposes, by accompanying dictionaries), the union of the words will define the text coordinates of the new common-information profile. When word counts are being mapped from their original databases to the new vector, the original TFIIDF weightings may be used, or new TF/IDF weightings may be created (using a dictionary constructed from all the databases' text taken together).

Once analysis has been performed, certain common-information profiles will be grouped together by their shared similarities, although the pseudonyms they represent may have been originally drawn from different databases. Such groups will represent links between different databases, and may be used for predictive purposes (see end of example).

6.65 Example of Cross-database Analysis

Suppose we have the following databases:

A. A travel agency keeps track of tickets sold, and vacation web pages browsed.
B. A bookstore keeps track of books sold, and stores an electronic version of the New York Times Review of Books.
C. A sporting-goods and clothing shop, keeps track of purchase items sold (which includes magazines, for which electronic text exists).

A certain airline wants to promote various vacation packages it has available, which include both European and Caribbean vacations, as well as singles and family packages. Although it has leased rights to databases A,B, and C, it turns out that no customer pseudonyms appear in more than one database at a time – in other words, there are no shared records.

A vacation expert is hired to create a common-information profile. He creates the following information vector:
(list of tropical countries, list of European countries, family score, list of sports, text)'

Note that the family score is a numerical value ranging from 1 (young singles) to 10 (many small children), and indicates what kind of person the customer is (a party-oriented student vs. a sedate father of three).

The expert the creates the following mappings:

A. Travel Agency. Link destinations of tickets sold to country fields (i.e., the number of trips to Germany by a customer would be placed in the Germany field of the common-information profile). Link sales of children's tickets, or requests for children's meals, to family score. Put web-page data into text field.

B. Bookstore. Link travel books' text to country lists. For all books purchased by a customer, map text from book reviews into text field.

C. Sporting-Goods store. Map warm-weather clothing (and swim gear) to tropical countries, ski gear to countries with skiing areas. Map sales of toys or children's clothing to high-value family scores, map revealing-bikini and student-discount sales to low-value family scores. Map text from magazines purchased by a customer to text field.

These mappings are then applied to each database, generating a full set of common information profiles. These are then clustered, forming groups that share commonalities.

The expert can now do several things with the results. First of all, he identifies the general "flavor" of each cluster (e.g., families with small children that enjoy winter, Europe, and skiing); the pseudonyms contained within each cluster can then be targeted for vacation packages suitable to their tastes. Secondly, the fact that pseudonyms from different databases have been clustered together allows the expert to plan cross-category marketing. If certain travel-book-buying parents have been grouped together with parents who bought their children swimsuits and scuba toys, it may be that they share a preference for family activities that take place in warm places or by the seashore. Hence, the book-buyers might be advertised various ocean-related sports goods appropriate for young families, and likewise the swimsuit-buyers might enjoy getting recommendations for travel books that describe tropical destinations that are especially fun for children. That is, if the goal is to cross-market items from A to customers in C, the most logical source of recoinmendations would be the people in A who have been grouped with the people in C.

SDI also allows dynamic and first-time personalization for a user that visits a vendor's site, on the basis of previous analysis and the user's profile.

6.66 Methods for Validation

The main result of using SDI is the creation of connections between different data points, linking vendors and customers, customers and recommendations. To a large degree, the overall success of an SDI analysis is the relevance of the connections that are inferred from the data. It is often the case that a certain amount of validation is required to determine which analytical approaches are the most successful, given that the analyst has had to choose a particular combination from a wide range of algorithms, data sets, levels of granularity, and parameter settings. The process of validation measures the relative success of a given project, and is used to guide the analyst through further iterations of tuning and adjustment so as to optimize the final results of the analysis.

There are two general approaches, not necessarily mutually exclusive, to validation: the first is fairly quantitative, the second relies more on human expertise and intuition.

- (1) Quantitative Approaches

1.1 Test Against a Validation Set

The goal of validation, in this context, is to measure how successfully SDI makes a prediction, most commonly a recommendation. Before a recommendation system can be used commercially (when it is exposed to actual customers), it is important to make sure that it is using the best possible combination of algorithms, input data, and parameter settings (e.g. TF/IDF tuning). If several different combinations are under consideration, there is a need to gauge the relative predictive accuracy of one approach over another. This can be accomplished by holding out part of the data set, training the recommendation system on the remainder, then evaluating the strength of the recommendations made for the hold-out set.

Suppose we are testing two possible settings for a system that recommends music. We make a copy of the customer purchase records and remove a single purchase at random from each customer – this slightly reduced copy will serve as our training set. We then allow the two rival systems to recommend musical albums for each customer, based on the information in the training set alone. Typically, these recommendations will take the form of a list of items with corresponding numbers that indicate the strength of each recommendation. The relative performance of a set of recommendations can then be gauged by looping across each customer, noting whether or not the system recommended the item that had been held out, and if so adding it to a running total. The system with the highest total can thus be judged the most effective, since it most strongly recommended items that the customers did, in fact, end up purchasing.

Because the result of this type of validation is a quantitative score, it is possible to automate the model selection process. Given a set of analytical approaches (each with its own array of parameter settings), it is possible to loop through the full parameter space (using a grid of evenly spaced numerical values, if needed, to reduce dimensionality), computing a validation score at each iteration. Those combinations of algorithms and parameter settings that demonstrate the best performance could be chosen as the top candidates for the final system configuration, since they do the best job at predicting customer behaviors.

1.2 Dynamic Approach

The problem with the hold-out approach to validation is that it isn't dynamic, since it doesn't reflect the impact that the recommendation system has on the customers once it is implemented, and may be based on data

that doesn't contain current trends. After all, it is better to predict what the customer will buy rather than what the customer has bought in the past.

A better approach is to run a controlled experiment against the actual customer base. First, the pool of customers is split at random into different segments. Next, each approach under consideration is used exclusively to make predictions for a given segment. Once the trial period is over, each system is given a score based on how valuable its recommendations turned out to be (this could be measured by total sales generated, for example, or by the number of times a customer made use of a recommendation).

- (2) Human Expert in the Loop

Although quantitative methods can automate the validation process to some degree, at the beginning of many projects there is so much raw input data available and so many decisions that have to be made about the analytical approach that an automated process would have to test a prohibitive number of combinations of data, algorithms, and parameter settings to get optimal results. In such cases, it is useful to employ a human expert who understands the psychology and nature of the particular domain being analyzed.

Such a person will have intuition about what is and isn't relevant for his domain. For example, a movie expert might be called in to work on a movie-recommendation system, for which an immense amount of input data is available. In choosing relevant fields for analysis, the expert's understanding of cinema would lead him to include the director's name and numbers of Oscars awarded, whereas the exact length (in minutes) of the movie would be, in his estimation, irrelevant and therefore excluded.

Once the analysis is complete and recommendations have been made, the expert's opinion (based on a qualitative understanding of the domain) can be used to guide which particular combination of settings, chosen from a list of candidates with detailed test outputs, should be used for the recommendation system.

- (3) Combined Use

There is certainly no reason why both approaches couldn't be used in combination. Many data sets include fields that are extremely noisy or simply irrelevant to a given problem; a human expert can be employed to pare the data set down to a reasonable size and dimensionality, using his domain expertise to create a data model reasonable for the proposed analysis. Next, automated methods can be used to fine-tune the parameter settings and to choose which subsets of the input data are the most useful. Finally, the human analysts called back to qualitatively evaluate the results of the fine-tuning, making the decision to either start a new iteration of the analysis, or to certify that the process is complete and ready for commercial application.

6.7 Conditions on Usage and Privacy

As previously described in this patent, various conditions can be placed on the way in which a set of data may be used (i.e., can the user make a personal copy of the dataset?), as well as on the privacy controls put in place. It might well be that a vendor is willing to share only a portion of his database, or that he will release only randomized aggregates in accordance with the level of privacy he has guaranteed his customers.

Although such restrictions could impact the content of the data analyzed by a vendor, as long as it is kept in an SDI-compliant format it can be analyzed by SDI's suite of tools.

However, the data that is stored in the central SDI sever still has tight usage restrictions, as placed on the data by the submitting vendor and the user that the data pertains to. For example, the user will have specified a use-of-data policy that could restrict the data to be used for only personalization purposes, or only solicitation purposes. In this case the user has connected to the site of a vendor, so this use-of-data restriction is irrelevant. The vendor will place additional restrictions on the data, for example the vendor might wish that the data that it submits is never used to personalize the

service offered by a direct competitor, with the same classifier label. When this is the case the central SDI server must not release any of this data to the vendor, even if this vendor is not a direct competitor, because there can be no guarantees that the vendor will not pass the data on to a direct competitor.

The data within the central SDI server is therefore in two main classes, data that can be released by SDI because the vendors that have submitted the data have placed no restrictions on the other vendors that can access the data, and data that can never be released by SDI because the vendors that submitted the data have placed restrictions on the other vendors that can access the data. We can still make use of the restricted data when providing profiling information to non-competitor vendors. Periodically the central SDI server performs collaborative filtering for each vendor, on the pseudonym records within its user base that contain information that relates to the business of that vendor.

For example, consider a vendor that sells compact discs. The vendor will have submitted profiling information about the each user (represented with a pseudonym) that it has done business with. Furthermore, the central SDI server might also contain additional profiling information that relates to the same pseudonyms, that has been submitted by other (non-competitor vendors). Finally, there may be records about other pseudonyms that have purchased music - either in compact disk format or an alternative format - that has been submitted by vendors that are happy to have the information used to help directly competing vendors.

The central SDI server can provide the vendor with the benefit of all of this data without actually releasing data to any of the vendors, by providing the results of the collaborative filtering analysis, restricted to the vendor's own data model. For example, given the connection certificate, that certifies that a user with pseudonym P has just connected to the site, the central SDI server can look up the profile of the pseudonym, and make appropriate product recommendations to the vendor - recommendations based on the analysis that has been performed using all of the data that it has available.

The structure of a typical data record contains a number of additional fields to represent profile-usage policies. (Pseudonym, data, L, R, P, S). The 'pseudonym' is the public key and ISP-level proxy server IP address relating to the pseudonym, the 'data' field is the profile information for the pseudonym - in general a sparse vector of numbers, L contains vendor classifiers that are excluded from the data, R is a {0,1} bit that indicates whether any of the data is randomized, P is a {0,1} bit that indicates whether the data can be used for personalization of service to the user when he/she visits a site under the pseudonym, and S is a {0,1} bit that indicates whether the data can be used for solicitations to the user with the pseudonym.

6.8 Location of Data and Algorithms

Although SDI might have controlled links to vendors' databases, the actual information might not physically reside within the SDI system. One could imagine a vendor requiring the joint analysis of data that includes highly proprietary information (kept for safety behind the company's firewall) with slightly less-critical information belonging to another vendor that is stored by SDI. In this situation, the vendor would have the ability, as with a lending library, to "check out" both the secondary data and relevant algorithms from SDI, and to use them at their own site. Thus, although both data and algorithms might reside either at vendors' home locations or within the SDI system itself, the general analysis will work transparently across these boundaries.

6.9 Other Issues

For security reasons, the contents of databases may be injected with a small amount of noise. This prevents database users from surreptitiously connecting database records to individual customers, yet maintains the quality of inferences made about the database in general.

Although such "noisy" records don't pose too much of a problem for those methods that make generalized inferences, it should be noted that recommendations made for individual customer vectors that have undergone such randomization will be less useful, since predictions are being made for a noisy target.

A final consideration is the reduction of the data vectors' dimensionality (which can be extremely high), since it is harder to make clean inferences about sparse data. There are many standard methods that can be used to achieve this, such as Principal Components Analysis.

Another approach is to adjust the granularity of the data, if at all possible. In a music store analysis, for example, there might be many more album titles that artists (since each artist can produce multiple albums). In such a case, purchases could be recorded by artist rather than by album, greatly reducing the dimension of the customer vectors' purchase space.

7. *Randomized Aggregates: Enhancing User Privacy*

Randomized aggregates provide a cheaper and more secure alternative to cryptographic techniques, such as secure-function-evaluation for providing information without compromising privacy. Even with pseudonyms it is possible that the identity of an individual can be revealed through revealing too much specific information about the transactionslprofile of an individual. This occurs when the information places a strong restriction on the set of users that could satisfy the revealed constraints. We enable vendors and users to report information about pseudonymized users in a randomized form that prevents this type of reverse-engineering to identify the user behind a pseudonym. We add randomization to data fields so that the user's privacy is protected, but also so that the data still allows accurate aggregation and collaborative filteringlmulti-attribute clustering analysis. Randomized data is also secure to computational attacks and the loss or theft of private keys-because we degrade the data, and make access to any one data item virtually useless.

This technique of randomized aggregates enhances user privacy guarantees, allows vendors to disclose useful information without violating user-privacy requests, servers to reduce system-wide reliance on cryptographic solutions. The basic idea is to add a small noise term to each field of a user's profile. Aggregate data can allow economic evaluation of data without access to the data, although aggregate data can have some inherent value in itself. Randomized aggregates enable: (1) personalization; (2) aggregate statistics; (3) protection against profiling.

In many situations it is sufficient to gather information about the activities of an entire demographic group rather than about any one individual. For example, a VCR dealer might be interested in the chance that a person buying a new television set will purchase a VCR within the next twelve months. However, information about whether any one individual who purchases a TV also goes on to purchase a VCR within a year should not need to be revealed in the process of computing this chance.

Traditional cryptographic methods are capable of solving the problem of revealing only aggregate information while concealing individual information. Methods exist for computing aggregates or other values from encrypted information without first decrypting this information. (Such methods are dealt with in an area of cryptography known as secure function evaluation.) However, the general-purpose nature of these methods makes them unnecessarily cumbersome for the limited problem at hand. In particular, the communication and computation requirements of these methods when applied to the problem of aggregation result in an unacceptable overhead on the system.

There is also an additional problem with such cryptographic techniques. Secret information can be compromised by successful computational attacks on the cryptographic scheme or by the loss/theft of private keys. Such problems are present in all uses of cryptography. Nevertheless cryptography is used where it is the best alternative. However, once again, for the limited problem at hand our solution is safe both from computational attacks as well as from private keys being compromised.

The above-described problems are solved by the use of a new technique that randomizes sensitive data. User decisions and other numeric fields of records are modified by the addition of a "noise" term that may be positive or negative. These noise terms are chosen by sampling from carefully chosen probability distributions. Different, independently chosen noise terms are used for each field of the record that needs to be perturbed by the addition of noise. The modified values of these numeric fields are then transmitted to an aggregating database. The net result is that when very few records have been aggregated the value of the numeric fields are likely to be quite unreliable. However the accuracy of these fields improves gradually as more and more records are accumulated.

As an example, consider the following situation. Suppose that we want to determine the average income level of people purchasing camcorders. In each record there could be a numeric field indicating the type of object that we are dealing

with. For example, assume that this field is "1" exactly when the object in question is a camcorder. Now suppose that each of these records has another field that contains income information. If this information has been perturbed by random noise then any individual record in the bucket has income information that is highly unreliable. However, when sufficiently many records are aggregated the income information becomes more accurate! (This is a consequence of a theorem in probability theory called the Law of Large Numbers.)

As another unrealistic but illuminating example, consider the following scenario: Suppose an organization receives monetary contributions from a number of individuals, who for reasons of privacy, wish to conceal the amount of their contributions. Suppose that the total of all contributions is to be public knowledge. The randomization scheme works as follows: When an individual contributes x dollars, she chooses a random number r according to a specified probability distribution (with mean 0) and sends x+r to the aggregate database. By allowing widely dispersed values of r, we can ensure that an eavesdropper who obtains the value of x+r has very little information about x. However, if a large number of individuals register their contributions in this manner, then the total of the values sent to the aggregate database will be very close to the true total contribution. The actual scheme and its variations are more complicated as they have to deal with multiple fields and more sophisticated attacks on privacy. But the basic idea described in this example will be used repeatedly in these schemes. In this example, the total of a set of values was required to be public.

7.1 Adding Noise to Fields

A record is a tuple of information containing various fields some of which are numeric, for example, describing the details of a commercial transaction. Noise or perturbation which refers to the random value added to individual numeric fields of records.

If the field is a continuous value, such as salary, then we can add a Normally distributed term with zero mean and a carefully selected standard deviation. The standard deviation is "tuned" to provide a good tradeoff between individual privacy and accuracy of aggregate analysis. For example, if an independent noise term is added to the salary field of a set of user profiles, and a vendor requests analysis of the mean salary over the set, then with a large noise term more users are required in order to generate an accurate average salary. In contrast, if only a small noise term is added to each salary field, and a third party with knowledge of the salaries of users requested data on each user, the third party would be able to match the users to the data fairly accurately, on the basis of the salary field-particularly for users with distinctive salaries (very low or very high).

If the field is one of a discrete set of items, such as the name of a CD that a user has purchased, then randomization proceeds slightly differently. In this case, one replaces the identifier with an identifier that is semantically close. For example, another CD of the same genre.

We can still perform correlation across fields with randomization, so long as the randomization does not destroy any trends between fields. Randomized data is marked as such within SDI, and labeled with the degree of degradation, so that SDI can be aware of the number of records to get relevant accuracy levels, and can report accuracy to customers.

We need to add noise to make data elements "close" to the accurate values. With discrete data, such as the name of an artist, "close" must be defined within the correct metric. The appropriate metric is such that a "close" value shares many of the same characteristics. For example, it is not appropriate to assign a close value on the basis of a shared last letter in the first name, but it is appropriate to assign a close value on the basis of an artist from the same genre of music. We call such clusters of artist names "semantic clusters", to imply that they have meaning in the domain. Semantic clustering that enables useful randomization of discrete field can be automated when goods are frequent purchase, high volume goods-where individuals purchase goods on multiple occasions, and more than one of the family of goods on a single occasion. High price, low volume goods, should be randomized on the basis of expert analysis (for example new cars, computers...) – where an expert can extract key features of a purchase, and represent the purchase generically using either a single prototype good, or one of a set of approximately equivalent goods.

We do not require that every vendor/user uses the same distribution for its additive noise term. The choice can be made autonomously. All that is required is that a vendor specifies the degree of randomization, so that SDI can be aware of the type of data that it is aggregating and selling.

We now begin an elaboration of the variations possible to the basic scheme and the assumptions and requirements under which each variation would be appropriate.

Records are associated with user pseudonyms, not user IDs. All fields can be perturbed with noise, except the pseudonym ID field, and any database query can be transformed into a query on "fuzzy" fields. Queries are dynamically constructed, and it is necessary to be sure that the result set contains enough data points so that a user's privacy is not compromised. Furthermore, we do not normally allow restrictions of pseudonym IDs.

The system can be configured and used in a number of different ways. Each transaction involves one or more parties. Most common transactions involve two parties, the vendor and the customer. It is assumed that the communication between these two parties involving the actual transaction is made secure by the use of a public-key cryptosystem or otherwise. This assumption will be valid in most electronic commerce schemes. Once the transaction is completed, the vendor or the customer is responsible for transmitting a record to the central SDI server. Again it is assumed that this communication is secure.

There are several mathematical issues to be resolved, foremost of which is the choice of probability distribution for noise. The system proposed here will be "tunable' for each application and it will be the application that will determine the appropriate noise distribution. Here we simply describe the issues involved in the choice of the noise distribution. An important requirement on the noise distribution is that it should have expected value 0 since we want the noisy total to converge to the true total as more and more entries are aggregated. The second consideration is the variance or standard deviation of this distribution. If the variance is made large (relative to the actual values involved) then each noisy value reveals little about the true value of a field in a record. Clearly this is desirable in order to preserve privacy. The drawback is that when the variance of the noise distribution is large, a large number of entries have to be aggregated before the sum of the noisy values approaches the sum of the true values. Thus there is a tradeoff between the level of privacy protection and the level of aggregation at which responses to queries become accurate. Once again, each application can determine the point on this trade-off curve most suitable for that application.

The choice of the degree of noise to add to a data field is a tradeoff between protecting the privacy of an individual and allowing robust profiling in the aggregate, and personalization at a pseudonym level. There is some information that can be gained from a "fuzzed" numeric value. Suppose that the noise terms are added from a Gaussian distribution, of unknown variance and zero mean. A third party can gain information on the likelihood that a fuzzed zero-one variable has a real value of one as follows: monitor a stream of fuzzed values and fit the most likely Gaussian noise distribution, or if the data is present in a database, just fit the most likely additive noise distribution. Then, given an individual fuzzed value x' and true value x, and an estimate of the statistics of the additive Gaussian noise distribution, it is trivial to compute the $Pr(X = 1 \mid X' = x')$. For example, if the noise distribution is distributed with mean 0 and variance 2, and the fuzzed value, x', is greater than 3, then it is more likely that the underlying true value, x, is 1 and not 0.

This is less of a problem when (a) the domain of the numeric attribute is large and/or (b) the variance of the noise term is large with respect to the domain and/or variance of the numeric attribute.

We can use a cryptographic technique to verify the distribution of noise that is added to data - and also to enable replay. A vendor must keep a record of the non-randomized data that is supplied. When generating a random perturbation, the vendor uses a one-way function f on the object X to generate a seed for a pseudo-random number generator. The pseudo-random number generator then generates a sequence of random numbers that are used to create the random perturbation from a well-defined algorithm. SDI can uses this technique to verify randomization, and audit a profile-updating agent. "Playback ability" - the ability to reconstruct the original record from a noisy version of that record is important for a number of purposes. An individual may want to obtain proof of a transaction for legal purposes and law enforcement agencies with appropriate warrants might want to examine original records. Each client-level SDI proxy S maintains a trapdoor function f (such as the RSA encryption-decryption function). When adding noise to a record the proxy uses the fixed fields of the record as argument x and computes the inverse of f when applied to x. (Note that third parties that do not have access to the trapdoor secret will not be able to compute this inverse function.) The proxy then uses this value as the seed for a pseudo-random number generator and uses the bits produced by the generator as the random bits used to produce the noise. With this scheme it is possible for the agent database to "playback" the noise perturbation and produce the original record from the noisy record. This playback scheme is optional and may be used in an application if the feature is desired.

7.2 Randomized Aggregates: A cheap alternative to Secure Function Evaluation

Randomized data also enables SDI to release data to third parties for analysis, but still apply its results to non-randomized data. The third parties cannot make use of data because they do not have the pseudonyms, and they cannot

identify any of the users. The models that they generate can be sold to SDI, matched with pseudonyms, and then rented to vendors.

There are two uses of randomized aggregates:

(a) Revealing aggregated data from a database does not reveal information that is specific to a particular user, and does not reveal information that is detailed enough to be useful to a vendor that wants to perform target advertising. It is information of the form "if you had the details, this information would be useful", and allows some calculation of probable economic value.

(b) When only aggregated data is required, then data can be transmitted securely in "fuzzed" form, and the aggregate date reported accurately given enough records. This is useful when the owner of the aggregated database is not trusted, or communication channels are not secure.

We need to check after a query that the data revealed does not compromise a user's privacy. A check at the time of submission of new data to the database is not sufficient because: (a) initially, no data is secure, even in randomized form - we can avoid a bootstrapping problem by checking the results of queries; (2) there are an infinite number of queries that one would need to anticipate at the time that data is checked in. We can use a statistical test to establish the minimum number of records that will ensure safe revelation of aggregate information, based on statistics/simulation from a real set of data, or statistics over a typical user population.

We would like vendors that hold useful information to be able to trade that information. This requires establishing that the information will be useful without revealing it. The classic solution to this would be Secure Function Evaluation. Secure function evaluation of the expected economic value of receiving information from one database to enhance the information present in another database is computationally expensive. Not only must the inputs for the function to be evaluated be transmitted over the network, that is the full representation profiles, but the functions must also be evaluated cryptographically.

An alternative to full secure function evaluation would be for an interested party to request aggregate information. Ideally this aggregate information would allow the economic advantage of the precise details of the new information to be estimated, without allowing the economic advantage to be immediately realized. The problem with using raw aggregated information, such as what fraction of users in your database that bought a TV also bought a VCR within a year, is that the requestor might already have information for many of the same users. What is required is that the requestor also transmits a constraint on the user ids that he/she are interested in.

Here is a simple scheme for computing the economic value of new information, for Vendor 1 that is interested in acquiring data from Vendor 2:

i) Vendor 1 sends Vendor 2 Query: What is the average income of the users in your database that have purchased a TV in the past 12 months, but are not in this list of users. (vendor 1 sends vendor 2 a list of users that it already has information about for this field).

ii) Vendor 2 can then send the aggregate information back.

iii) Should vendor 1 be interested in receiving some form of access to information about the aforementioned set of users, then negotiation over terms of contract can proceed.

iv) If the negotiation is successful, then finally vendor 2 can send vendor 1 the information, authenticated in some way, and vendor 1 can check the accuracy of the aggregate information that was reported in step (ii).

v) Payment for the information is made contingent on the accuracy of the information provided in step (ii).

Another interaction could look like this:

i) Vendor 1 says "Do you have attribute X for these users". <sends list of user ids>

ii) Vendor 2 reports "Yes, for 89% of them".

iii) Vendor 1 makes a deal with vendor 2, and they agree payment contingent that the data that vendor 1 receives is consistent with the aggregate information provided in step (ii), and verifiable.

Finally, third party certification is necessary to guarantee that vendor 2 is providing truthful data for the unknown fields of the records requested by vendor 1. This aggregate data approach will be useful when aggregate data is sufficient to perform an economic analysis on the utility of new information.

Aggregate information does have value. Consider a vendor that has performed some analysis on sparse data from his/her own database and is seeking to support his/her conclusions by performing a similar analysis on the sparse data

of another vendor. This analysis could proceed with aggregate information alone, and have value in confirming trends observed in the vendor's own data. The vendor could use the analysis of data from a similar vendor to draw implications about his/her own database of user information. SDI must perform secure evaluation on information in this case. The two different methodologies for the determining the value of information prior to disclosure of that information are:

Aggregate information has no direct value to a requesting vendor, but can be used to assess whether the raw information will be useful.

In this scenario it would seem preferable to avoid the complexity of requiring a trusted third party to perform the evaluation. Instead the aggregate information can be provided directly to a requesting vendor, allowing the vendor to perform his/her own analysis. Aggregated information is always sufficient, because the only way that a vendor could use non-aggregated but randomized information would be to aggregate the information himself. We would never need to transmit randomized records to the requesting vendor, because randomized information is only useful in its aggregated form anyway. For example, the vendor might request aggregate information on the users for which it already has information within its own database, or aggregate information only for new customers. Should the aggregate analysis by the vendor indicate that the information held by the disclosing vendor is valuable, then mutually beneficial terms of disclosure can be agreed between the requesting vendor and the disclosing vendor, and the disclosing vendor can send the requesting vendor that information in a non-randomized form.

Aggregate information has direct value to a requesting vendor, even without the raw data from which it is derived (as described above).

In this case we will require a trusted third party to evaluate the potential value of information held by Vendor B for Vendor A. This will require the specification by vendor A of constraints, and possibly a sophisticated algorithm by which to evaluate the utility of the data in the aggregating database. Note that the variables used within such an evaluation function must be limited to aggregate information (means and sums) because that is the only accurate information that can be derived by the trusted third party from the aggregating database.

The trusted third party responds to the requesting vendor, and the requesting and disclosing vendors are then free to pursue negotiation in order to establish mutually agreeable terms of disclosure. If the data that is finally requested by the requesting vendor is aggregate data, then that can be disclosed directly by the third party. Otherwise disclosure must be between the vendors themselves.

7.3 Randomized Aggregates: Further Uses

Mutually agreeable terms of disclosure of that data can be negotiated between the requesting and disclosing vendors. If the data is aggregate form, it is disclosed directly by the third party vendor otherwise, the non-randomized data is disclosed by the appropriate vendor to the requesting vendor. The aggregate database may be used for any of the following purposes:

(1) The use of explicit queries submitted by a given vendor to determine whether another vendor (whose aggregated data resides in the secure database) possesses data which matches the criteria of the external request (typically in conjunction with certain stated constraints, as discussed below).

(2) A comparison evaluation of the vendor's aggregated data to determine whether and if so, which portions of the database (possessed by which vendors) contain data which is of potential relevance to the vendor's database.

(3) The generation of "virtual customer lists" which are determined to be of value to the vendor (through the evaluation process of either #1 or #2 above). The pseudonym proxy server may then be used to target the desired users matching the requested criteria of the vendor based upon terms agreeable to the requestor and data providing vendor based upon a per-impression or per-transaction model (through use of a changeable pseudonym).

(4) If sophisticated formulae and/or esoteric algorithms are used where evaluation/processing of a significant quantity of a given user's data is required, it may be preferable to confidentially convey these formulae and algorithms to the third party. The outputs from pseudonymous users or the list of users which achieve certain predefined outputs may then be conveyed to the vendor.

(5) Disclosure of randomized data by the randomized database to a vendor in order to develop formulae, algorithms and models.

As discussed further below, the value of the data to the vendor may be estimated by:

1) (if the prospective disclosed data is users), the number of users matching the relevancy criteria.
(2) (if the prospective disclosed data is statistical in nature), the degree of statistical confidence associated with it.
(3) (if available), the effective measurable degree of benefit resulting from previously disclosed similar data (if applicable, with similar statistical confidence) as to that of the present data to be disclosed.

We can also determine the number of other "undifferentiatably similar" users that the user's identity can be masked by such that they are unidentifiable from the user. This is easily estimated by comparing the randomized data set to a set of real data, developing a randomized version thereof and observing the degree of confidence in being able to predict the true identity of the randomized user profile of a given "user"(from the randomized data) as confirmed by the actual data set. The present system may also notify the user if/when the number of identical users falls below a minimum predefined threshold to ensure this requirement

8. *Techniques of Secure Function Evaluation*

While randomized aggregates address the user privacy concerns associated with enabling vendors to gain full access to user profiles for modeling and in order to develop algorithms that fully leverage the new information that can be collected on-line, there is another potential user privacy concern. In order to actually deliver ** personalized service ** to an individual user, it may not be sufficient to use that individual's randomized profile, or an individual might be reluctant to reveal the randomized profile that is associated with her pseudonym, if it does indeed carry useful information.

A technical solution to this problem is offered through secure function evaluation, where two parties can evaluate a function based on distributed information without gaining any information about the inputs other than what is revealed by the outcome of the evaluation. The result of this method is the secure automatic evaluation of the user's profile using the algorithms and tools originally derived from randomized aggregates. The output is assured to be accurate by guaranteeing that both the user data and the vendor's imported function are secured to tampering by either party. The client-level SDI proxy server maintains a user's accurate profile data, while submitting randomized profile information to the central SDI server.

In one variation, the secure function evaluation can be used to determine whether there is data in the databases of other vendors that will be useful to enhance the user and object profiles of a vendor. If another vendor's data can be enhanced by the type of data on the vendor's site (and amenable terms are agreed upon), the vendor can then provide the useful portion of his/her data. Secure function evaluation allows the evaluation of data without access to the data, so that a fair contract can be agreed between the parties. The evaluation and exchange (and associated negotiation) of the data, occurs on a trusted server or the SDI proxy server where the relevant data resides. Because this analysis, data collection, compilation and modeling procedure occurs securely and under the control of the third party operator or intermediary operating the distributed trusted server network (e.g. SDI) the integrity of the data as well as secure enforcement of the privacy policies of the parties which possess access privileges to use that data are assured.

Practically speaking however, some of the user profile data (e.g. direct mail orders) will be directly collected by the vendor (this is inevitably true for any purchase of physical products whether directly using a store card or mail order with the exception of anonymous physical mail described below) and/or some users may not require that their data be protected by a pseudonymous proxy server. Accordingly, the secure function evaluation may also be used on behalf of the vendor to assess what portions of the data on the user's database are of potential value to the vendor. This potentially could involve such database matching activities as a simple direct comparison of fields of data between the user and vendor database in order to determine missing entries and/or fields on the vendor database to more sophisticated inference of data fields and/or leveraging additional relevant statistics for supplementing a sparse matrix for purposes of predictive user modeling purposes. This provides the user with a high degree of privacy, as no portion of the user profile needs to be revealed to the vendor.

The vendor criteria for selecting both users and appropriate targeted ads (based upon their user profiles) from a profile database can be performed entirely autonomously by the use of the present secure function evaluation method. Once the secure function evaluation has identified and qualified the relevant data and also measured the predicted benefits to the vendor, the vendor may possibly choose to negotiate terms for full purchase of the user data which is specifically relevant to that vendor. This data can be down-loaded to the vendor's site to directly analyze his/her user behavior

statistics. The requesting vendor's "representation profile" consists of the complete collection of target object profiles to be refined.

The first step is searching for and identifying data within remote vendor databases which are likely to be the most ideal candidates for cross advertising with that vendor. Secondly, the database(s) which contain valuable data are evaluated for the estimated degree of potential commercial benefit the data could provide. Thirdly, this value estimation and the data privacy policies of the disclosing vendor (based upon hislher interest in providing data to the requesting vendor) form the basis for terms, if any, for the data transaction).

One example of two different types of synergistic data bases include vendor databases and a distributed user database (e.g. on the proxy server). Large customer interaction and transaction object profiles on the vendor's site are likely to be much more robust that those collected by a user side profiling system distributed over a multi-server network, or even multiple networks. From the vendors perspective the breadth of domains containing products that can be potentially cross correlated from data on the user's database is very valuable. Vendors can hope to reach customers who have visited other similar sites and sites which tend to be visited by similar users. Thus, the ad network is able to utilize the user database (containing virtually all of the data about what sites and target objects a user accesses) in order to effectively identify other vendors whose products and content are most frequently accessed by the most similar users(and types thereof), suggesting the best target customers for each other's advertisements.

The system can contact the remote vendor requesting secure access to hislher data through secure function evaluation in order to determine the estimated degree of statistical benefit. The vendor may choose to disclose data in randomized form in order to enable the ad network to improve its data model while avoiding disclosure of data for purposes of targeting hislher customers. If the vendor lacks installed cluster code in his database, the system automatically installs the code as part of the secure function evaluation. In order to save overhead of re-installation the next time that a request is made to access the vendors database or to update the present model the system requests the user for permanent installation of the secure function evaluation and associated analysis code. This enables the automatic updating of a distributed cluster model in accordance with dynamically changing information and user behavior patterns. The cluster code is actually loaded onto the remote site, the analysis by the secure function evaluation may also determine the estimated potential benefit achievable to the remote vendor (or user) associated with accepting the request for hislher data. This may be achieved by comparing previous similar situations in which disclosure of data occurred, i.e. the marginal measurable commercial benefits resulting from previous data sets with similar degrees of statistical sparseness including the quantity of pre-existing and newly introduced data within a similar domain or cluster to the present one. The degree of estimated commercial benefit resulting from disclosure of the requested data may be then calculated. A separate estimation is then made if the remote vendor agrees to disclose hislher data (in non-randomized form) for purposes of targeting ads to customers on hislher site (or email list).

Depending upon the existing degree of statistical confidence which is estimated between each target customer of the remote vendor and the target objects of the advertising vendor, for the most part, it is likely that utilizing techniques which improve the statistical confidence between users and target objects of these metrically similar vendors will yield significant commercial benefits by improving the relevancy of the matches. It is also possible for a data privacy policy to be submitted in conjunction with the release of any data. The vendor may desire that certain conditions of release be associated with all or each portion of the data, in the form of explicit restrictions of usage which may be assigned and tagged upon its release by the vendor. Adherence to these restrictions is the responsibility of SDI.

These data privacy policy restrictions may indicate which other vendors/vendor types (if any) may receive the advertising benefits resulting from that data, what portions of that data can be utilized for the benefit of each vendor/type. Also the pricing structure may vary in accordance with the type of data released, and who the vendors are who will ultimately benefit from its incorporation. Additionally, the usage of the data may also be subject to controls set forth by the contributing vendor. For example, applying different restrictions to different portions of hislher data, restricting its modeling benefits to hislher own company exclusively, limiting the benefits to non-competitors or certain vendors, limiting it's usage to only certain advertising types (or purposes), restricting its usage to individual impressions or sales (on a per impression or per transaction basis). Data may only be used on a per-impression or per-transaction basis.

Because of the restrictions which vendors may place upon the usage (what portion of their data, to whom and/or for what purpose), re-clustering the data model separately according to various degrees of disclosed statistical data can be computationally expensive. We can use the same cluster model for portions of the model with common disclosure policies. Each portion of the vendor's resulting data model with the same disclosure policy is called a "common disclosure unit" or "CDU". Delineating CDUs within the data model involves continuous monitoring and identifying

those portions of the data model (attributes and associated metrics). The CDUs, whether common access, limited common access or totally "private", operate seamlessly with CDUs that are subject to less restrictive constraints.

It is also worthwhile to mention that more simple merging techniques may be used for enhancing remote marketing databases by adding or filling in data fields to a sparse data model. This may be explicit data from matching fields, or inferred data, e.g. additional products or attributes which statistically correlate with certain other products or attributes or other user attributes. The system can notify vendors of the presence of potentially useful data within the user database.

Secure function evaluation can also be used to retrieve target objects from secure private databases (or Web sites) that would otherwise be inaccessible. SDI can identify whether there are any metrically similar target objects in the remote database that match the target objects currently retrieved.

We use the cryptographic technique of secure function evaluation to evaluate the encrypted profiles of users that pseudonymously access a vendor's site, but prevent the vendors from accessing the profiles directly. The profile-analysis package is tuned for an individual vendor by analysis of the vendor's data model in SDI. The analysis is provided to the vendor on a restricted basis by requiring that (a) the proxy servers request a new encrypted profile for each user every day; (b) the vendors request a new secure function evaluation module every day (or a new key), so that vendors cannot use the analysis for more days than they pay. Secure function evaluation computes results from encrypted information without decrypting that information.

9. *Leveraging Existing Standards*

The architectural framework outline above can be implemented with a number of existing technical methods. In this section we outline one possible approach, that uses the Extensible Markup Language (XML) to encode web information with meta-tags (that can represent profile information). The Java virtual machine shipped with current Internet browsers can be used to run personalization code, that determines dynamically the information that should be presented to a user.

We can use XML to embed the profiling information for products that are offered by vendors, and information that is offered by information providers, directly into the page itself - with semantic labels that allow client-side processing with a Java engine. The Java Virtual machine, implemented on the browser of a user's client machine, takes as input the XML-tagged page of a vendor, and the locally stored profile information that pertains to the user's current pseudonym, and generates personalized content to display on a monitor local to the user.

Compare this architecture to a traditional client-server based solution, where the server produces personalized content for the user, content that is pushed to that user. Such a system architecture requires that the server has direct and explicit access to profile information about the user, information that can then be exchanged with other third parties and is out of control of the user. Secure Data Interchange collects and distributes information consistent with user- and vendor-defined policies. Information can be explicit data, including transactional information that is collected by parties that are involved in a transaction—such as the product purchased, and demographic information (e.g. gender, zip code, occupation). Information can also be implicit data, that includes click stream data that logs the information that a user requests and views, and the time-sequence of hyper-links that a user follows as he/she browses across multiple web sites. Profile information is embedded within a web page as metadata, that is data about data - machine readable information that informs an intelligent agent (such as an SDI-enabled browser) about the data that is included in a web document. The extensible Markup Language (XML) proposal of the Worldwide Web Consortium working group on SGML provides an ideal standard for representing such information [XML].

XML allows meta-content to be included with documents, machine-readable information that enables documents to be processed by client software. Augmenting web documents with structured information in SDI enables clients to perform user personalization - pushing computation to clients, and allowing greater control over user-profiles because profiles do not need to be released from clients. XML can represent rich data structures, and that allows a grammar to be defined for information that allows data to be automatically verified for correctness [SGML].

The ability to embed data within web pages allows client-side processing of information. By embedding profile and location information directly within a web document we can alleviate the bandwidth and computational bottlenecks that can occur at a centralized profile server if profiles are fetched on-the-fly when web pages are downloaded by clients.

The origin server (supported by the vendor) requests periodic profile updates from the central SDI server. This duplication of information enables the profile and the page contents to be provided directly from a vendor's server.

There are some potential drawbacks of this approach: (1) the profile information associated with a web page and target objects can be out-of-date; (2) the profile information is available to all clients and proxy servers, not just those that are SDI-enabled; (3) the profile information can be altered. We suggest technical solutions to each of these problems below.

9.1 Periodic Update of Web Page Profile Information

The central SDI server provides profiling information to vendors that subscribe the SDI. A vendor sends a 'Request Profile Update' message to the SDI server, to instruct the SDI server to send new profile information. The SDI server responds with a 'Profile Update' message, that contains updated profile information, generated from explicit and implicit data that it receives from vendors and users in accordance with privacy policies. The request-response mechanism can be implemented using the standard HTTP Post/Response mechanism in conjunction with XML message types. For example, the server 'Request Profile Update' message can be represented in XML as:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:Request> http://www.some_vendor.com </SDI:Request>
```

and the SDI server 'Profile Update' message can be represented in XML as:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:Update>
      <SDI:Profile>
            <SDI:Item> (1231, 0.453) </SDI:Item>
            <SDI:Item> (1041, 0.034) </SDI:Item>
      </SDI:Profile>
</SDI:Update>
```

An illustrative Document Data Type (DTD) for an SDI:Profile element type is presented in the next section. The XML messages are included in the body of standard HTTP Post/Response messages. We limit the performance degradation caused by out-of-date profile information that is stored within web pages of on-line vendors by associating "out-of-date" time stamps with the profiles that are provided by the central SDI server. This mechanism is similar to the "expiration time" tag of a Netscape Cookie message. The frequency with which profile updates need to occur will depend on the speed with which profile information changes. The "out-of-date" time stamp can be included as an additional element in an SDI:Update message.

Vendors request new profile updates when the current profile information is out-of-date, and more frequently if required (although we allow for a per-update charge). This is a "pull" model for profile-updates. An alternative architecture for profile updates is a "push" model, where SDI periodically sends new profiles to be added to the web pages on a vendor's server. The "pull" model is our preferred model because the responsibility for maintaining current profile information is decentralized, resting with the vendors in the system.

The system as outlined above can be implemented within the current HyperText Transfer Protocol (HTTP), as a sequence of challenge/response pairs between clients and servers. The HTTP Post/Response mechanism allows clients

and servers to exchange data, and this data can be an instance of an XML Document Type, within the body of a HTTP message. The HTTP protocol is the underlying mechanism, with SDI messages contained in the body of the HTTP Post and HTTP Response as XML documents. In one variation of SDI the profile of a user is maintained on the user's client, and partitioned into separate profiles for each pseudonym that a user chooses to maintain. Personalization of products and services (product types, prices, etc.) is performed at the client, through the execution of trusted code that is embedded as a Java applet or as JavaScript within the web document of a vendor. In this way a vendor never receives access to the profile of a user, but is nevertheless able to personalize its response to users, even when a user first visits a site (on the basis of the profile for a user from his/her previous online transactions). Profiles for the target objects of a vendor that enable appropriate objects (representing particular products, or news stories for example) to be presented to a user are embedded as XML data within the vendor's web document.

In another variation of SDI personalization is not performed at the client, but either at the ISP-level SDI proxy server or the vendor's server. The location and other profile information that relates to a user are pushed to the ISP-level proxy or vendor server when a user requests a web page. In the same way as XML allows profile information about web sites and vendor products to be associated with a web document, and profile information to be provided from the central SDI server to a vendor, XML can be used to encode a user's profile. The system of SDI allows for profile and location information to be randomized slightly (and even anonymized) to protect the identity of a user, for example when an ISP-level proxy is not trusted.

9.2 Example: A Possible XML Representation of a User Profile

The World Wide Web Consortium (W3C) SGML working group developed XML (extensible markup language) to provide an open and extensible grammar for structured data [XML]. An XML document has an associated schema definition to enable an XML-enabled browser to validate the structure of XML data automatically. A Schema in XML is called a Document Type Definition (DTD), and defines the names of tags, their structure, and their content model. XML allows the DTD for an XML file to be identified through a Universal Resource Indicator [URI] in the header of the file (see below). XML also allows URIs for mobile code resources to be referenced, in order to enable a client to process embedded XML data. An XML document must be well formed, and in order to be well formed the tags must form a tree structure. In addition, the DTD allows the structure of an XML document (an instance) to be validated against a particular schema. Senders and receivers must only send valid SDI files. Each SDI message is a valid XML document.

We provide an example XML instance and part of a Document Type Definition for use within the system of SDI. Profile information, as generated automatically through collaborative filtering techniques (for example, see issued US Patent #5,754,939) can be represented as a list of attribute-value pairs within an XML document. An attribute is defined by a numeric code, and the value defines the weight of the attribute. For example:

```
<?XML version = "1.0"?>
<?xml:namespace ns = "http://www.w3.org/OPS/OPS" prefix = "OPS" ?>
<?xml:namespace ns = "http://www.sdi.com" prefix = "SDI" ?>
<!doc>
<SDI:ProfileData>
   <SDI:Location>
        <SDI:Geocode> 12321561 </SDI:Geocode>
        <SDI:DigiMap> http://www.digimap/?12321561 </SDI:DigiMap>
        <OPS:Zip> 19103 <SDI:/Zip>
   <SDI:/Location>
   <OPS:Demographic>
      <OPS:Gender> F </OPS:Gender>
      <OPS:Age> 26 </OPS:Age>
      <OPS:Income> 50000-75000 </OPS:Income>
   </OPS:Demographic>
   <SDI:ID>
        <SDI:Pseudonym> P12543 </SDI:Pseudonym>
```

```
            <SDI:PublicKey> 12453246129421 </SDI:PublicKey>
        </SDI:ID>
        <SDI:Profile>
           <SDI:Profile-item> (1242, 0.546) </SDI:Profile-item>
           <SDI:Profile-item> (56,   0.045) </SDI:Profile-item>
        </SDI:Profile>
     </SDI:ProfileData>
```

The Document Type Definitions for this document are specified in the header, and include URIs to a DTD of the Open Profiling Proposal of the W3C, and also a DTD of the Secure Data Interchange. The OPS DTD is used to boot strap the SDI DTD, providing tags for common profile information, such as 'Gender', 'Age', 'Incomc', etc. The section of the SDI Document Type Definition that is used in the above XML fragment is presented below. It makes reference to tags defined in the OPS DTD, and the RDF (Resource Description Framework), a W3C proposal to standardize the structure of DTDs for XML documents. XML Name spaces [NS] provide a method for unambiguously identifying the semantics and conventions governing the particular use of property-types by uniquely identifying the governing authority of the vocabulary, for example OPS and SDI in the example above. The URI for a schema can contain a human and machine-readable description of an XML schema.

```
    <!ELEMENT SDI:ProfileData (SDI:Location?, OPS:Demographic?, SDI:ID?,
    SDI:Profile?) >
    <!ELEMENT SDI:Location (SDI:Geocode?, SDI:DigiMap, OPS:Zip?, OPS:Address?) >
    <!ELEMENT SDI:ID (OPS:Name?, SDI:PublicKey?, SDI:Pseudonym?) >
    <!ELEMENT Profile RDF:list<SDI:Profile-item>  >
    <!ELEMENT SDI:Geocode #PCDATA >
    <!ELEMENT SDI:Digimap #URI >
    <!ELEMENT SDI:PubicKey #PCDATA >
    <!ELEMENT SDI:Pseudonym #PCDATA >
    <!ELEMENT SDI:Profile-item (SDI:Attribute-ID, SDI:Attribute-value) >
    <!ELEMENT SDI:Attribute-ID #PCDATA >
    <!ELEMENT SDI:Attribte-value #PCDATA >
```

The tag '#PCDATA' is used here to represent numeric or textual information, '#URI' declares that an instance of element 'SDI:Digimap' must be a valid URI pointer.

9.3 Maintaining the Integrity and Security of Embedded Profile Information

The privacy of information in transit between servers and clients can be assured through standard end-to-end cryptographic solutions that establish a secure session prior to any data exchange, such as Secure Sockets Layer (SSL) that uses X.509 certificates and is supported by current browser technology.

In order to prevent the possibility of individual users being bribed by vendors to disclose target object profile data which reflects this type of information, users should not be provided access to directly decrypt the metatags for these portions of the target objects profile data, but rather this decryption and release of profile data should be performed securely in conjunction with the functions of the profile processing (profile matching module) upon the client level proxy server rather this decryption and release of profile.

In addition, we prevent unauthorized access of embedded profile information through the encryption of the metadata that is represented within the XML structure of a web page. Profile information can be encrypted using a hierarchy of keys, so that different levels of access to the information may be provided according to the access levels of users and vendors. All users that request web pages from SDI-enabled vendors, whether or not the user is a member of SDI receive the same profile information. We provide encrypted profiles to vendors in the 'Profile Update' messages from SDI to vendor servers, so that: (a) unauthorized agents cannot tamper with the profiles; (b) the profiles cannot be read by unauthorized agents.

The SDI system supplies a private key to trusted SDI client software, that enables only SDI-enabled clients to access profile information, and only access that information to the extent permitted by privacy policies of users and vendors. Different levels of encryption enforce multiple levels of access. Periodically the key pairs are changed to prevent

extended attempts at cryptographic attacks. The SDI system uploads the key that provides the correct level of access for a user to a user's client, once terms of access and profile management have been agreed. A client can only access embedded information once enabled with a relevant key. Finally, profile information is signed with a digital certificate, to prevent third parties from tampering with profiles for commercial gain.

9.4 Related World Wide Web Consortium Proposals

The proposal for a Meta Content Framework (MCF) suggests a particular structure for the description language for web pages, to enable schema to be shared and re-used [MCF]. This proposal is incorporated into the W3C Resource Description Format standard [RDF]. The proposal for an Open Profiling Standard [OPS] describes a system for profile exchange between two parties, building on XML and MIME standards. The W3C SGML working group has defined XML to provide an open, extensible grammar for structured data. The proposal on privacy and profiling on the Web [PRIVACY] extends the vCard [VC] schema for electronic business cards to include profile information, and suggests that profile information can be stored and managed locally, with Client-server exchange of personal information as required (using the HTTP challenge/response mechanism).

The Resource Description Framework (RDF) enables the encoding, exchange, and reuse of structured metadata. RDF is an application of XML, with additional constraints to allow for DTDs to be published, and interchangeability across different communities. The ability to standardize the declaration of vocabularies will encourage the reuse and extension of semantics among different information communities [Mil98]. RDF is a W3C proposed standard for defining the architecture necessary for supporting web metadata. RDF is an application of XML that imposes structural constraints to provide unambiguous methods of expressing semantics for the consistent encoding, exchange, and machine processing of metadata. RDF additionally provides means for publishing both a human-readable and a machine-processible vocabularies designed to encourage the exchange, use and extension of metadata semantics among disparate information communities.

10. Support for E-Commerce Functionality

10.1 Generation of Mailing Lists

We can use the same profile information that provides focused/personalized service to users that hit a site that they have not visited before to form well-targeted mailing lists for vendors. The Secure Data Interchange can form mailing lists in a number of different ways.

First, consider a vendor that wishes to send targeted mail to some of its own user-base. When users connect to a site they indicate whether or not they are willing to receive electronic mail, and provide a "mail certificate" to a vendor if they are happy to receive mail. The Secure Data Interchange can proceed as follows:

(a) perform analysis for the vendor to determine an appropriate set of users to receive the solicitation, based on the information that the vendor provides about what it intends to market, and provide the list of pseudonyms to the vendor for mailing; (b) perform the same analysis, but also forward the communication to the users directly.

Now, consider a vendor that wishes to target new users, represented with different pseudonyms. Users indicate whether the information that a vendor submits about his/her transactions may be used for solicitations, and furthermore vendors indicate the set of business interests that can receive the benefit of information that is submitted to the central SDI server. The SDI server can continue by performing analysis on the relevant subset of the permitted class of data records that pertains to the product or service that the vendor wishes to model, and generate a list of appropriate pseudonyms. Finally, the SDI server can sell the pseudonyms to the vendor outright, together with a certificate that the vendor can send mail to the pseudonyms, or the SDI server can retain control by sending the mail on behalf of the vendor.

Provide vendors with virtual mailing lists that can be mailed to via the proxy server only, i.e. these customers should be solicited based on our analysis. (could even give summary info., without revealing details about users). Furthermore, do not even reveal data that corresponds to a pseudonym to a vendor because the vendor then has that information about me when I am on his site.

A central data warehouse also enables vendors to identify new potential customers. This process is broken down into a number of steps:

(i) The vendor assesses the value of the information present in the secure data interchange. This computation is performed securely either by revealing randomized aggregates to the vendor to enable its own local analysis, or by allowing the vendor to check data and algorithms into the secure data interchange site for analysis.

(ii) The vendor selects criteria for mailing unsolicited advertisements, and agrees on a pricing model. In this case per-impression pricing is the most obvious pricing model, as it is difficult to monitor when a user responds to unsolicited mail per-transaction pricing is difficult. The user could be motivated to do this should the Secure Data Interchange promise future returns for recording a successful solicitation with the database.

(iii) Either the data list is released to the vendor for its use, if this is within the selling vendor's data policy, or the data interchange sends mailings on behalf of the purchasing vendor.

10.2 Physical Mail

(a) Vendor to User

Figure 13 shows a flow chart for the process of sending physical mail from a vendor to a pseudonymous user.

A vendor must hold a "physical mail certificate" to be able to send mail (packages, letters) to a user under a pseudonym. The certificate is similar to the "electronic mail certificate", in that it is signed by the private key of the user's pseudonym, and indicates that the vendor with public key P*V can send mail to the user (under the pseudonym).

Each user has a trusted physical address authority, just as it has a trusted electronic mail authority (the second-level proxy server), that maintains the physical mailing address for each pseudonym. When a vendor has a letter X to mail to user with public key PKP, the vendor generates a unique ID for the package, IDX, and sends the ID code and the physical mail certificate to the trusted physical address authority of the user.

The physical address authority receives the certificate, S((PKP, PK*V, SEND–MAIL), SKP), that indicates that the vendor is authorized to send mail to the pseudonym, and the packages identify code, signed by the vendor to certify that the vendor holds the secret key that matches the public key in the physical mail certificate.

The vendor then passes the letter X and the signed ID code to a trusted mailer, that supports pseudonymous mailing, and has been certified by the central SDI server as such. The trusted mailer then provides the signed ID code to the physical address authority, signed with the private key of the trusted mailer. The physical address authority verifies that the trusted mailer is a valid service, and releases the real address of the user to the mailer. The mailer now has the letter X that the vendor wants to send to the user with pseudonym P, and the physical mailing address of the user - and the package can be mailed. At not time did the vendor determine the true mailing address of the user, unless it works in collusion with the trusted mailer, but the trusted mailer is certified by SDI, and also audited by the chosen physical address authority of the user. The address authority will only release addresses to reputable pseudonymous physical mail agents.

We can operate physical mailing lists in the same way, and gain additional security by never releasing the pseudonyms or the mailing addresses to the vendor that has requested the targeted solicitations. We can use a technique that is similar to the technique that we used for virtual mailing lists. The vendor describes its solicitation to the central Secure Data Interchange, which leverages as much data as possible (without violating the privacy policies of any of the users or vendors that are represented within the data). The central SDI server generates a list of suitable pseudonyms, and then provides a series of unique codes to the vendor, that the vendor can supply to its chosen pseudonymous mailer with the material that is to be mailed. The central SDI server also provides the appropriate address authorities with authorization to release the physical mail addresses to the mailer when presented with the IDs. Notice that at no stage did the vendor have the pseudonyms or the mailing addresses. The parties all have only as much information as is necessary - the vendor needs someway to identify its packages to the pseudonymous mailer. The mailer needs an identifier to present to the address authority, and receives the addresses. The address authority just needs to know what addresses to release and to which third party.

(b) User to Vendor mail

The Secure Data Interchange system also provides a mechanism for users to send physical mail to vendors that are registered with SDI with pseudonymous return addresses. In particular, when a user sends mail to a vendor, the first-level proxy server provides a tool that: (1) Computes/Looks-up the appropriate pseudonym for the user with this vendor. (2) Generates a unique ID, and submits a signed message to the central SDI-server, where the message relates the pseudonym, the vendor, and the ID. (3) Provides the unique ID to the user.

The user writes the unique ID on the envelope, and mails it to the vendor. Should the vendor wish to reply to the user, then the vendor can take the envelope to a pseudonymous mailer, and request that the envelope be mailed appropriately. The pseudonymous mailer verifies the identity of the vendor, and then submits the ID, together with the vendor's signature, and its own signature, to the physical address authority that is maintained by SDI. SDI releases the address to the mailer that can then return the mail.

10.3 Pseudonymous Payment Mechanisms

The Secure Data Interchange architecture must be able to support all the standard electronic commerce functions that we take for granted, but while maintaining pseudonymity for users and following privacy policies. There are various different solutions to this problem.

(a) Anonymous Credit Card Payment

The second-level proxy server can maintain information on the user's credit card information, and perform the following transaction. Whenever a user makes a purchase from a vendor, the user provides the vendor with authorization to bill $x to hislher credit card account, but anonymously - through the Secure Data Interchange as a middleman. The user generates a unique number, Y, and signs a "right to payment" message, M=($x, PKP, PKV, Y), that gives the vendor the right to make a claim for payment of $x from the Secure Data Interchange. The first-level proxy server registers the unique number Y with the second-level proxy server to ensure that the vendor does not spend the money twice, and provides the proxy server with authorization to charge $x to his/her credit card when the request for payment is presented.

When the vendor submits its "right to payment" and proof of identity to the second-level proxy server the proxy server first runs the charge through the user's credit card, and if that clears, runs the charge from the vendor through the account of SDI (which could also be a credit card, or could be operated as electronic cash or some other mechanism for payment).

This "anonymous credit card" payment method has the following properties:

1. The user's credit card pays $x, but does not know who receives the money except that it is going to the Secure Data Interchange.
2. The vendor receives payment for $x, but does not know the user's credit card information, or the user's identity.
3. The Secure Data Interchange incurs no financial risk because it receives payment from the user before making payment to the vendor, although there could still be problems if the user complains about the quality of the good for example.

This protocol is simpler than full cryptographic anonymous credit card mechanisms because the SDI acts as a rusted third party to both the user and the vendor. << provide literature references >>

(b) Electronic Cash

Electronic cash is anonymous, just like physical cash. The user purchases electronic cash from an electronic bank, presenting blinded notes, so that the bank has no record of the note numbers that it issues to the user. For example, the user generates a new note number, X, and has the bank sign a blinded copy with its $10 signature, S(B(X), SKBANK$10). Then the user, or the first-level proxy for the user, removes the blinding factor, and can use the

electronic cash as tender. Whenever the note changes hands the recipient needs to check with the bank that it has not yet been spent, because notes are easily copied, but not forged.

Electronic cash has the same useful properties as anonymous credit cards, although it is perhaps a little more exotic. In particular, notice that the bank does not know to whom, or for what, payment has been made. and the vendor does not know which user made the payment - it just receives the payment. We have minimized the amount of information exchange that takes place between the various parties in the system.

11. Models of Data Release

Data can be released with associated "terms of disclosure", that define: (a) the price or "exchange rate" of data; and (b) the time and usage parameters of access to the data. Possible time and usage parameters include: data may be used indefinitely, a periodically renewable usage contract, a number of impressions contract, a number of transactions contract, as a one-time permanent or temporary exchange, as a time limited privilege, or subject to certain conditions.

Data can be priced according to its accuracy and content. For example, aggregate data might be sold more cheaply than detailed data, and we can sell the right to access data dynamically (on demand), on a per-impression basis, or up-front. Some vendors might wish to purchase the rights to utilize the data. The vendor might also be interested in renting data (for example, when the value of data is uncertain), and purchasing the data outright if it proves valuable.

The central data warehouse of the Secure Data Interchange can operate as a competitive market place for data and information, with the possibility of facilitating the trade of information between vendors that have synergies in their data sets, perhaps with SDI acting as a valuation device to provide information about the likely benefits of an exchange. The form and terms of data exchange will be those that provide greatest benefit to the contributing vendors.

In general, vendors will most likely be interested in purchasing the access to profile information about users that hit their web pages dynamically, on a per-impression or even per-transaction basis. This removes much of the uncertainty from the vendor about the coverage and relevance of data that is provided.

The methodology that we promote within the basic SDI architecture, where users themselves maintain control over their profiles also provides users with ownership of that information, and enables users to maintain ownership while providing per-impression access to vendors, who receive the results of using the profile to personalize an interaction but not the profile itself. The profile need never be released by the user.

Per-transaction pricing, possible in on-line applications of secure data interchange, offers powerful new modes of business. For example, it is possible to monitor the number of "click-throughs" achieved on a particular banner ad,? and charge the requesting vendor on a per-transaction basis. This provides a self-enforcing structure to a contract. It is in the best interest of a providing vendor, and the secure data interchange, to provide accurate data to ,enable proper focusing of ads, and also to provide good data analysis tools, because the success of the advertising campaign determines the revenue that they receive. It is as though the providing vendor is working on an "on-commission" basis.

Ideally we would like to sell data to a vendor at the value of that data to the vendor. This should be done in a way that prevents the vendor from selling the data on to another vendor on the black market. One solution to this problem is to "rent" the data to a vendor, but not give a purchasing vendor outright control of the data. The secure data interchange can be used to provide access to the data. For example, instead of physically transferring data to the data warehouse of a purchasing vendor one could provide a mail-distribution service for that vendor at the data interchange. The vendor could specify constraints on the users that are to receive a particular solicitation, give the interchange "Mailer" the advert, and then request that the interchange deliver the solicitation.

The interchange could also provide a vendor with the right to run analysis at the interchange, without actually down-loading the data.

12. Variations on the Basic SDI Architecture

There may be commercial contexts in which an SDI service can be established where there is already in place a pre-existing trust relationship between multiple vendors and a third party. Such third parties are inherently motivated to provide services to enhance advertising and e-commerce for their existing and potential customers. These third parties may include, for example, web hosts or e-commerce service providers (ESPS) which often have hundreds or thousands

of sites which they host, Web portals, information and commerce service manufacturers, advertising and affiliate network services and data analysis and business intelligence tool providers (which includes the business to business application).

A third party may wish to implement an SDI which operates separately and independently from the central SDI service. Alternatively, some of these third parties may install an SDI server on their customer information server. The server may be integrated into an existing advertising service which they operate and maintain, in which case the vendor receives an appropriate fee for data which is exchanged between his/her existing customers, and a reduced fee (which may be split with the central SDI service) for data which is exchanged by/between a member of his/her SDI service and vendors who are members of the central SDI service but not of his/her local SDI service.

12.1 An open SDI system

An ISP level proxy server can contain the user profile generation module, profile processing module, user profile interest summary generation module and target object generation module which operate in distributed manner. This enables an ISP to independently implement the core functionality of the system without the cooperation of information vendors (Web sites) or their operators (Web hosts). The modules in third-party SDI servers can share information with the modules in network vendor servers. This flexible architecture enables ISPs to implement SDI and when available also the complete data sets available from the information vendors.

SDI can allow third parties to operate their own secure advertising and/or electronic commerce-based product syndication affiliate network (for all customers). The users that are also subscribed to SDI, can be given highly personalized information for each site or for the network of hosted sites (which could involve an interface which provides site to site links as a "virtual mall"), and a menu interface to these sites which includes the 2 or 3 dimensional personalized menuing features and personalized search facilities as disclosed in the parent description (a "personalized portal").

12.2 A Closed SDI System for a Syndicated Network

The Web host (or more generally a vendor, a provider and/or operator of server functionality to a variety of information vendors), may also be interested in operating his/her own closed version of SDI. The main SDI server for the closed system can be located on the network vendor servers, or it may reside upon the information vendors servers (as it is operated by that local Web host). For example, an affiliate ad network (including a web host acting in such capacity) could upon installing SDI onto their network enable and enforce the wishes and desires of advertisers (and particularly) sites which are advertised upon with regards to what types of sites and advertisers (respectively) they allow or disallow for purposes of standard or affiliate advertising, in accordance with the methods herein disclosed. The general implementation for determining which this general application for using collective user feedback to determine relevant site links was described in the parent issued patent). In this case, end users who are subscribed to SDI would receive personalized affiliate links (including product level recommendations for on-site purchases) which have been pooled and profiled at the main SDI server from all SDI vendors (in distributed fashion) and matched with the user.

12.3 Interoperability Between Local SDI Services

With interoperable (local) SDI services, we can also facilitate the secure enforcement of data sharing policies and transfer of transaction fees between these local SDI services E.g., by/between aggregations of ad networks, syndication networks and Web hosts operating virtual portals and advertising/syndication networks with personalization as its primary capability.

In each of these primary example domains, the server operator is financially motivated to sell the SDI services to his/her sites because the transaction based model is used, and the server operator receives the commission on each transaction (or click through) occurring within his/her network of sites. However, if the server operator also integrates his/her local SDI service into the main SDI service (to share user lists and impressions and/or space to advertise to these target users), s/he can receive a commission (in conjunction with each vendor transacted with) for each advertisement placement or syndicated transaction to or from his/her network.

We can also allow the local server operator to split the transaction fee (normally received from the main SDI service), thus "referral fee" for both the referred customer and the referral of customers (through the placement of outside ads or products on one of hislher sites) or other means of targeting hislher site's existing customers.

Reduced overhead resulting from economies of scale which may likely result in incentives to the local operation, e.g., free installation and operation of hislher local main SDI server by the main SDI service, i.e., as the operational overhead would be cost justified by the shared transaction fees of customer referrals and advertising space coming back to the main SDI service.

This architecture also may be useful and is ideally suited for cross vendor product advertising as through an ad network or product syndication network using affiliate links. In addition to the user profile generation module, a target object profile generation module should also reside across the network vendor servers such that it is possible to generate target object profiles for target objects on network vendor servers. Alternatively, user profiles and target object profiles are downloaded to the client level proxy which performs collaborative filtering tasks as the user browses from site to site.

In both of these cases, the main SDI server can receive user profile data generated from the user profile generation module located on the ISP-level proxy, and target object profiles generated from the target-object profile generation modules located on the various multiple information vendor servers.

13. Dynamic/Real-time Secure Data Interchange

The user-centric SDI model allows users to provide personal information on a carefully controlled basis to vendors and other users. Furthermore, vendors can implement rules that personalize the information, products, and services provided to users—on the basis of personal information that they receive from users directly, or have acquired about users.

As an extension to this model, we also allow users, vendors, and other third parties to associate "meta-information" with other users and vendors. This information might be a user's opinion about hislher interaction with another user, an annotation that relates to a particular web page, or information about a physical object, for example how to get to the top of a tower. The system of SDI enhances the value of this information by providing a secure environment where users can also associate their own profile with the meta-information that they "leave". This allows collaborative filtering techniques to generate appropriate meta-information about an object (user, physical object, vendor, web page, etc.) that will be useful to a particular user—given that user's own profile.

We define "virtual tags" as any piece of information about an object (physical or virtual). The information may be authored by any party, but annotated accordingly. For example, the appropriate virtual tag provided by a user about hislher-self is the pseudonymous profile for that user, -- and with SDI only the user his/her-self can gain access to the profile (either directly through editing, or indirectly through continuing transactions).

It is useful to implement a "reputation system" within such a virtual community. Initially users (under pseudonyms) have no reputation, and their opinion does not count for much, but after every positive interaction (as defined by other parties in an interaction), the "reputation" of a user can increase. (see the Kasbah system, MIT). This reputation system is appropriate to a pseudonymous environment. Notice that gaining negative reputations is not useful when users can simply change identities.

In one variation we can "block" certain users from providing information, when those users have negative reputations. Clearly, collaborative filtering or other data mining techniques could usefully allow for reputations when weighting information about an object.

13.1 Autonomous Exchange of Information

Client-level SDI proxies can act as autonomous agents in an architectural variation of SDI, where the "client-level proxy" is co-located with a (physically) mobile user, for example on a palm-held computer or head-up display.

In a "match-making" application domain, the goal of user-agents is to find other user-agents with desired profiles/synergies, and arrange person-person meetings, or business-business meetings/agreements. There are two

parties in any exchange of information, and although information exchange will be bidirectional, it is useful to talk about a "requestor" and a "requestee". The key provision provided within SDI is that:

1. Requestors and requestees can communicate anonymously, without revealing (even pseudonymous) identities.
2. A requestor cannot access profile information about a requestee unless authorized to do so (implicitly or explicitly) by the requestee.

Implicit authorization occurs when a requestor can present certificates to verify that it has required attributes to access particular information. Explicit authorization occurs when a requestee provides direct authorization to a particular part of a user's pseudonym.

Essentially there is bidirectional information filtering: the requestor agent will only present certain information to the user, information that is relevant; and the requestee will only provide information when a request is judged to be legitimate. Information exchange between agents occurs as part of a multi-step negotiation, until both parties can agree on terms for either a physical meeting (or execution of a deal), or further pseudonymous exchange of information or cooperation.

In this case, using methods taught in the co-pending patent << LEIA >> user-agents can identify other agents that are "close" through an anonymous matching market, where agents provide their location and a (one-time) identity for contact. The market informs agents when other agents are close. In fact, this "anonymous matching market" can be extended, and agents can provide more than just location information.

A user-agent might also broadcast a "persistent" query over the agent-network, for example requesting response from agents with a particular set of attributes, and providing some information. Decisions about what information to exchange are made on the basis of both static and dynamic profile attributes, e.g. standard (historic) profile information, current behavior, current location, recent activity, and credentials that can be presented/denied. LEIA style-behavior attributes can be used to automatically decide on the relevance of new virtual tag information. A requestee might also demand certain credentials to indicate the lack of negative reputation marks, for example that an interaction with the user has never received a bad rating. Perhaps a third-party could be used to determine whether the user's know each other (eg www.sixdegrees.com)

We can extend information disclosure to include communications between users and other parties. A user-agent might decide to make another agent privy to communication, on the basis of the context of communication, content, location of parties, profile of the third-party, etc.

When a requestee denies a request for information, it may instead provide criteria for data releases. A requestor can respond with a different information request, or a subset of required credentials. Finally, the agents might agree on terms of negotiation and conditions can be anonymously fixed.

Negotiation might FAIL, in the case of missing rules in a rule-set, or a special case that has not been anticipated. In this case direct user-user interaction might be necessary, although this can be executed anonymously via a real-time anonymizing service.

There are (at least 5 levels) of information disclosure:

1. indicate to another user interest
2. release of profile information
3. disclose communication
4. add an individual to a current correspondence session
5. schedule a meeting/strike a deal

The end-result of information exchange could be an agreement to calendar a meeting for some future time and place; and absolute, or pseudonymous revelation of identity.

Initially, an implementation of the data-release policies might allow only manual definitions. However, after an initial "beta testing" phase, a data mining suite could be used to cluster users and generate exemplar data release and data request policies. A system can provide default settings for users, and recommend setting based on users with similar profiles. The user can further fine-tune the rules.

Finally, automatic feedback techniques can be useful to adjust rules, for example—when a user is especially receptive to particular type of introduction then make such introductions more likely in the future. An intelligent interface system

might also suggest refinements to the rules, to automatically cover "patches" where the user currently controls interactions. (see the Microsoft Lumiere project).

Note 1: must next describe user profile attributes and rating criteria, e.g. selective access to group ratings and annotations by navigating their attributes. Also automatic generation and comparison ratings between items via rapid profiling.

Note2: Real time (including spoken) communications (either user to N-users or user to user) within a local proximity where the users are "pre-qualified" (e.g. knowledgeable in that particular field, educated, etc.)

13.15 Credentials: Examples and Typical Uses

In accordance with the parent issued patent, various credential issuers are provided for issuing standard and resolution credentials to individuals. Thus certain entities may be entrusted with "legitimate authority" to validate and submit credentials which are issued to the appropriate individuals. Resolution credentials are provided to prove the absence of a quality attribute or behavior (which is often of a negative nature) relating to an individual and is submitted by a third party and typically must be issued on a periodic basis in order to maintain currency. If a resolution credential is not issued (or not renewed) an adjudicating third party is provided which has access rights to both of the parties is provided to resolve resulting disputes (from the subject user). The present invention describes how credentials can be issued to users pseudonymously.

A few simple examples of resolution credentials which may be of interest to users (credentials which users may commonly request as a precondition to requesting or accepting requests to be introduced or initiate communication with an outside unknown third party) include:

1) (for business associations) are in good business standing, e.g., have not attempted to defraud other users in the course of common business practices. Or maintain sufficient finds in one's account to perform business activities (as represented by the user).
2) (for business interactions or social interactions) are in good standing with the law.
3) are considered "safe" individuals,e. g., have not been rightly accused or criminally convicted of violent acts, (e.g. a store clerk may wish to be made aware when an individual walks into a store which cannot present a resolution credential for not having been convicted of retail theft or robbery.
4) have not been accused by other individuals of inappropriate or antisocial behavior.

Some standard credentials which may be of interest to many users, and which may (as with resolution credentials) be incorporated with the standard settings of the user's data request policy as herein described. A few examples are cited (among countless potential others): profession, awards, honors, alma mater, e. g., Harvard graduate, doctorate degree, etc.

There are a variety of rules which a user's data disclosure policy and data request policy may contain, to control what if any attributes are released, and what credentials are required. A data request policy may state a rule for explicitly notifying the user if a particular resolution credential (e.g., indicative of a serious problem or concern)

cannot be presented in response to the
user's disclosure request.

Some users may not wish to disclose specific information about themselves via these standard credentials but instead certain "extracted" more general information may be provided about themselves. For example, instead of a "Harvard grad or Ph.D." there may be, for example, credentials indicating "intellectual" or "prominent intellectual". Or instead of indicating an individual's wealth or value of assets, the credential may indicate;"wealthy" or "very wealthy" (typically, depending upon user's wishes this latter credential should also be withheld during initial introductions or subject to some fairly stringent conditional criteria from the other party) and instead replaced with an even more general credential e.g., "prominent" or "influential citizen"). Similarly, an individual's exact profession or scope of work may not be fully disclosed initially but rather a more general definition of his/her profession or perhaps the general field initially in which the user works.

Another example of a credential of potential interest may include the profiles of users which a certain individual associates with or is acquainted with. The ability of a third party to gain access to this information, however, is

conditional upon the data release policy of that associate's or acquaintance's data (e.g. it could be affected by what is the profile of the common acquaintance to whom that user would be disclosed as an associate as well as, importantly, the profile of the prospective disclosee.) In one variation, the system may simply identify the fact that there are common associates and acquaintances between the two individuals. Again that associate's or acquaintance's data release policy may further control even detection of this fact. It may instead also notify one of the parties of this fact, but request that it not be disclosed to the other party.

In accordance with the parent patent application, rules may be learned regarding certain things that a user does (as in ascribing these rules for which messages to send to whom or what user profiles and under what circumstances/events surrounding the target user). Thus, his/her agent may begin to suggest certain future actions which could be performed in the future upon user approval or even automatically. If the user has had no previous interaction at all with the system, it may identify which other users of the system the present user is most similar, and recommend initial rules. Additional textual attributes can also be leveraged to provide extra criteria, and data mining techniques used to generate more appropriate rules.

Another category of user credentials include features that may be inferred implicitly by location/time data captured by LEIA. Such information may reveal a user's likely behavior and activities. These inferences, however, are unavoidably somewhat speculative and inconclusive, thus cannot be substantiated on a valid basis for issuing credentials. The data may be useful in suggesting the present context and circumstances surrounding a user.

Additionally, the communications which the user may be presently involved in i.e., the content profile of his/her spoken dialogue and/or other "on line communications" may be used and combined with location/time patterns in order to further infer the circumstances, behavior, and present temporal interest of a user and/or third party for purposes of employing the user's data disclosure and data request policies.

** DP—but how can we do this when users are only pseudonymously identified? How can LEIA map physical user IDS to pseudonymous IDs?

There are many potential application contexts in which this present architectural framework could be usefully deployed. For example, it would be extremely useful for users who wish to be re-identified or re-contacted after original introduction -- with unique pseudonyms. A user might wish to withhold all identifying information from his/her acquaintances and associates, except when required for further information exchange. The user can also apply the techniques of randomized aggregates to minimize the chance that one friend or associate will be able to correlate that unique pseudonym with the same individual. A new portion of the use's profile could be combined by users with different information about the user or more importantly, a private piece of information under a pseudonym could potentially be associated with hislher identity. The user agent might use a statistical algorithm to identify a threshold of how much data can be released (including what types of data and to what degree it must be randomized), in order to prevent the user's identity from being compromised.
Within a location enhanced context, unless the prescribed range of "proximity" to the user is quite large, securely protecting the user's identity from malicious third party collusion (for purposes of combing unique pseudonyms and/or exchanging data that has been released and entrusted to them) is a harder problem. The system could (most obviously) assume data exchange between the parties will occur and limit the combined disclosure to only that of the most data restricted user in a given location/time domain. The system could alternatively, perhaps "space apart" the number of users within a given locationltime context who can access more "restricted" user data (of course the problem goes away if all the disclosees have similar disclosure restrictions by that user).

We allow initial information exchange to be anonymous, such that information that is released as preconditions for release of further information is not useful. Similarly, so long as initial encounters are anonymous there is no need to withold information about them from the user.

** DP—this is a cleaner solution than an "ISP-level user agent".

Credentials can allow users to identify other users that may pose a threat. This identification may be provided vis-a-vie resolution credentials and/or rating (by third parties). e.g. a user has not engaged in any serious criminal activity, physically harmed another person, or interacted with other individuals who are unable to produce these resolution credentials. Other credentials may specify the nature of an infringement, and its context and severity (e.g. what was the context of a physical assault? Was it performed during a bar brawl, against a friend, a boss, an elderly person, a child, a family member – or at work? In this case, the user agent may, for example, bring to the attention of a prospective employer that the user could not present a credential indicating that they had not previously harmed or threatened a

former employer. Was it minor or severe? Also, if such individuals (lacking, for example, resolution credential proving the absence of having committed armed robbery) are (or come) within a certain proximity of a user, the user may wish to program hislher user agent to notify the user. The same would, of course, apply to a store clerk regarding customers of this sort or to baggage security personnel at an airport. Or, highway patrollers may be interested (e.g., on certain stretches of highway) in being made aware of vehicles and their locations whose agents are unable to provide a resolution credential proving the absence of a drug conviction.

** DP—in general we cannot expect other users to be "broadcasting" negative credentials. The most that we can expect is that a user with negative credentials can not provide (false) positive credentials, and assume that the lack of positive credentials implies the presence of a negative credential.

** DP—think about infringements in one pseudonym can prevent a user acquiring a positive credential under another pseudonym

In a more benign though similar example application to that above, users who are "common individuals" could use the above techniques (through similar co-operation/collusion) to "single out" other individuals who may be of a "high profile" nature (e.g., famous or wealthy) who may otherwise strongly prefer to remain incognito, wherein potential user to user communications could result in a significant invasion of privacy.

In another application (in accordance with the auto insurance risk determination methods described in co-pending patent application entitled "Applications for Location Enhanced Information Architecture"), an on-board computing device within a user's automobile could identify another automobile lacking, for example, a resolution credential for safe driving, i.e. the on-board user agent continuously polls agents in other cars for a "safe driving" credential, and if it fails to receive such a credential it issues a warning to the user. As an extension, this location data could be converted into a dynamic 2-D rendering upon the user's windshield (using heads up display technology) in order to thus superimpose a persistent flagging or highlighting of that particular automobile from the driver's visual perspective. Pedestrians could also receive instant notification.

We can avoid user-identificationlprivacy violations within such a physical system by fuzzing the information provided to a user, for example indicating only a general location, or hiding the individual within other individuals. When highly sensitive information is disclosed, it is important that we can protect the real identity of the user.

Some examples may include: health problems, like HIV status, neuroses, sexual abuse as a child, family history of violence, alcholism or emotional disturbance. However, a prospective partner might wish to receive credentials of a lack of these (negative) qualities before pursuing (non-anonymous) contact, As is described in LEIA, a roaming cellular connection, or GPS, is not essential for providing a user identifier. For example, optically-based biometric identification techniques such as iris scanning or combined iris/facial identification techniques may be used among other potential inputs as well. Users will be reluctant to realease locationltime data, even anonymously, when suspicisous behavior can be inferred—probably subjectively. Of course, we allow the user to control this data release within his/her privacy policy.

Indeed, if a "significant" threshold of suspicious behavior is exhibited and detected,the information may be accessed by law enforcement officials, through seizure of the decryption key for that data (which includes hislher physical location information) and any additional profile data which is considered of immediate critical relevance to the suspect (or prospective) infraction. Such cryptographic techniques for key seizure from a key escrow are well covered in the literature. If behavior is inferred from LEIA information, then the threshold should allow for statistical errors. There may also be certain circumstances in which key seizure may be required after the fact (at some time in the future).

For example, if/when certain even moderately "suspicious" behavior patterns are detected, it may be possible for the SDI data warehouse to preserve a comprehensive record of that information (and perhaps the record of that user which precedes and follows that period of interest). Thus preserving evidence which may later prove useful in contributing evidence towards a conviction, acquittal, e.g., proving that a user was not at a particular locationltime. A record containing more detailed segments of a user with a proven negative or questionable history may be preserved and general location/time features may be abstracted for the remaining portions of the record (thus compressing the record substantially). This may be performed for regular individuals as well, thus retaining key relevant features while discarding the majority of the record which is irrelevant or redundant.

A very important caveat underlying the above architectural framework is that until such time as users become personally "wired" (through miniature computing devices and/or wearable computing devices), the above applications involving the use of resolution credentials within the context of a location enhanced (physical) environment will be hard to implement practically. Automobiles may be an exception,as may be technology which enforces the disclosure of the physical presence of a user agent (resident in a device) to other user agents within the networked environment.

Finally, we need to protect users from prejudice, on the grounds of religious beliefs, political affiliations, sexuality, and membership of certain organizations. A user's inability to provide resolutions credentials could result in a significant threat of the fundamental elements of their individual freedom and civil rights. Thus credentials proving the absence of such affiliations/memberships or professions should again be banned from use. Thus, an important condition should be provided as a part of the system's design criteria enforcing the recommendation that users must collectively avoid the use and request of such resolution credentials as part of the user profiles

13.17 Additional Applications of Virtual Tags

The following is a list of additional types of application specific information which may, for general or specific applications, be usefully deployed through the use of virtual tags.

(a) Buyer and seller information - as detailed in the parent issued patent, specific details of what buyers and sellers may be looking to buy or sell respectively may be used to suggest the basis for a potential commercial transaction. The transaction may be large (but not necessarily so, e. g., real-estate, private investment in a small business or public stock). If a physical or on-line interaction with the other party is warranted (e. g., for larger commercial transactions), as is suggested later is in the present description, users may identify other users which fonn the most relevant "match" with their interest. At this point the agents can check for credentials, and then either communicate or calendar a meeting. Similarly, the agents may find the
"best" match of users who happened to be physically proximal to the user at that particular time, or at some future time(s)/location(s) which is mutually compatible (similar applications are suggested for matching sales persons with prospective clients, identifying experts to work (individually or collaboratively) on a particular project or problem, to answer a question of an appropriate specialized nature to their area of expert knowledge.) The parent issued patent suggests at a general level these commercial applications. An additional feature described therein involves the use of a decision tree called "Rapid profiling" which can be used in the present context to identify from the most common needs of buyers and "goods" of sellers in general and the known profile data about each buyer and seller individually, a list of questions for each party which most briefly and efficiently determines the complete buyerlseller profile of each party individually.

(b) Medical information, such as medical conditions, medical history, active prescriptions, drug reactions, family history, possibly even genetic pre-dispositions (from a genetic profile). Medical insurance information may also be potentially useful for a prospective qualified accessor to be able to readily access in case of an emergency.

(c) Social Interests Profile Information -- The parent issued patent also suggests the present application at a general level. For a dating application, users may be matched on the basis of their common interestslpreferences and perhaps on the basis of certain information reflecting personality, social or cultural behavior/affinities or psychological attributes. On the other hand, for purposes of meeting casual acquaintances, users may be interested in another user who shares the above characteristics as well as someone who has recently shared similar experiences and/or personal challenges.

(d) Physical location information -- Users or advertisers could, for example: a) Query a pseudonymous user database to access profiles that are in close physical proximity and match certain criteria, e.g. live in a certain geographical region, had recently attended a meeting or event (or is planning to attend a particular event) had recently communicated with a friend or associate. In another variation, a user could for example, submit a query pertaining to every user in a particular physical space, e. g., a room, hotel or convention center, e. g., identify all users present here who attended Internet World, 1995.

(f) Professional Information/Qualifications - As in the application of matching buyers and sellers, a description of a user's needs or situation with relation to various professional services may be provided as additional data about the

user. Examples may include: (as above) medical data, professional or business history (as well as legal history) which may be of interest to law firms, accounting firms or various business consultants. Personal , family or emotional difficulties may be of interest to psychologists or family counsellors. Again, users may submit this information as a query for prospective matches, or they may be pseudonymous queries or automatically matched in accordance with criteria specified by the professional. The issued parent patent application also lists additional applications which could as well be relevant within the usage context of virtual tags.

Referral Information - Situations frequently arise in a variety of contexts of human interaction (whether social or professional) in which a user may wish to refer the user they are in contact with to another individual. Often this occurs in a professional services context which a user has a particular need/or other characteristics which make himlher an appropriate match for the services provided by the other party. Or in a business context, often a user will forward a business contact or associate to another colleague who is deemed more appropriate for the particular context and/or scope of business. Likewise, in a personal or social context users may sometimes meet two or more individuals which they observe or perceive share common interests, goals or beliefs or perhaps possess complementary capabilities, knowledge, or characteristics. In each of the above scenarios, virtual tags may provide substantial benefits.

For example, the referring user could forward the relevant portion of the profile and identified need of the user to the referring party whose user agent may determine the acceptability of the request and/or the priority with which a communication or meeting could be scheduled (e.g., as could be automatically arranged by/between the two party' calendaring agents). If the referring party's agent is unable to make a decision or priority assessment for scheduling purposes) on behalf of the user, the agent could instead try to contact the individual himlherself for assistance (and statistical feedback to the system's data model). In order for these types of referrals to be performed efficiently, the area of expertise required can be specified, and provisions can be made about the type of referrals that a professional will accept.

(h) Employer/Employee Information - Users who are seeking employment (actively or unofficially/passively and employers who are seeking employees for specific responsibilities may benefit under the current scheme. An employer may post a description as part of hislher virtual tag (and that associated with his/her company). His/her employees may also have provided ratings and/or annotations which are further descriptive of his/her personality, leadership/management style and skills, work environment which s/he promotes and overall quality. Previous employees for that position (who may also have either provided information about themselves as deemed appropriate) or may also have provided such information as well as pertaining to the position. If not s/he may allow himlherself to be contacted by the prospective candidate (e.g., in exchange for a fee).

Access Privileges Information - Users in an organization are frequently given privileged access to certain files within a corporate intranet but not others. Though there are many ways of profiling users according to their level of access privileges to information, the following example is considered: Based upon the position (e.g., responsibilities and tenure with the organization), users may be "classified" into groups according to different levels of access to confidential information. Virtual tags may be used to extend the capability by providing for immediate disclosure of a user's information access privileges to another employee in real-time and in a physical context. Similarly, it is conceivable in either a professional or non-professional environment that users may wish to maintain "secrets" between each other. Access or restriction to/from a secret may be provided in accordance with a particular attribute of a user (e.g., departmental secret, a family secret, personal secret) or by the individual identity of the user, e.g., upon contact/communication is privileged or not to a given secret per the restrictions by its originator).

Additional rules could be provided, so that if certain pre-defined events occur, the criteria for release of information associated with the secret are changed. Alternatively, this "information" triggered by an "event" could instead be a message. For example, if a user reads or accesses certain information, meets with a certain colleague or friend send message X. This message could be (for e.g.) a request to perform some task relating to part of that information, a reminder to address certain issue(s) while chatting with the colleague etc. or, per the request of an individual's employer or colleague if a given individual (a sales person) meets with user X send himlher message Y (which may refer to a previous encounter, experience or fact slhe should know pertaining to user X and which may have bearing upon their conversation or professional interaction. (This message could as well be sent by user X as a reminder notification to himlherself at the time of the encounter).

In another application, the user's access privileges may be used for granting him/her access to restricted physical areas, (thus, the virtual tag effectively may behave like an "electronic door key"). A variation of the technique may be used for granting access to professional meetings, where information access privileges of users must match the anticipated confidentiality parameters for the scheduled meeting. Another application may include the ability to automatically enable access or restrict access, based on payment of fees, and whether or not an individual is a representative or partner of a competing company.

(i) Contextual Information -- If user X is a prospective customer, and performed certain on-line or physical behavior that suggests an interest in a product/service, then the advertiser's agent can automatically send an ad (subject to the user's privacy policy). Another condition which could trigger either a notification to the advertiser or the message (automatically) could be temperature or weather conditions (which may affect people's purchasing behavior as well as often driving/traveling activities which affects the agent's scheduling and reminder'slfollow-up with participants on existing scheduled meetings.

13.2 Example: Match buyers to sellers

For example, in a mobile sales-force, the system of SDI can first generate an ideal list of prospects, and then help salespeople target products and offers to individuals. We can provide information to salespeople about users, according to the profile of a salesperson (and reputation), and a user's personal terms for data-disclosure. Similarly, a system of SDI in conjunction with the methods taught in co-pending patent << LEIA >> allow automatic detection of salespeople close to users (via an anonymous location market). The market allows matches to be made, but does not reveal anything about a user that the user does not authorize.

For example, in a marketing network, with a commission-based sales force. User profiles can also be used to determine user-responses to offers and products (see the methods in patent application << system for personalized ..>>) User profiles are generated from interactions with other vendors, salespeople, and current activities/behaviors. SDI allows profiles to be built from extended interactions across multiple vendors, so long as the user authorizes the same pseudonym for each vendor.

Similarly, users can themselves use seller profiles, to decide whether or not to interact with a seller, and vendors can user seller profiles to decide how to allocate new customer prospects to sellers. The profile of a sales-person may show correlations between product sell-rate and the type of product, type of user, that the sales-person interacts with. Initially seller profiles may not be very well related to sales-performance, but instead based on general SDI-style profiling, and wider (eg professional) credentials. Later, as a seller gains experience, profiling can be based on a sales-person's track record (and this will subsume other information).

As an extended example, we can also consider a system for handling tasks—e.g. a call center, or a dynamic work-flow system. As new jobs (or calls) arrive into a system, the jobs are automatically routed to the appropriate expertslmachineslsales-people according to the ability of the person to perform the task. The ability is measured by a match between the profile of the person and the type of task, for example on historical performance/feedback, and on other relevant attributes, and also through collaborative filtering techniques. The goal of the system is to allocate tasks in the most efficient manner, across a system of experts (or machines).

The problem might also be informational: e.g. find an expert on ancient American civilization for purposes of writing an article, or answering a specific question. Relevant information might include the expert's resume, and the expert's knowledge expertise profile developed from his/her activities in responding to previous queries.

Level of expertise might also include the size of projects performed within a particular specialized area, and relevant education qualifications.

Example: business-business introductions. Another application domain for privacy-protected match-making, where users are anonymous until an agreement is struck is business to business introductions. For example, it might be useful to automatically identify synergies between businesses (e.g. in infrastructure, technology, or product) -- for the purposes of pursuing an advantageous strategic relationship. If the meeting is between two employees of competing companies, then the system of match-making could also ensure that a meeting is predicated on a particular task that does not cause conflicts with their respective companies.

We can also use a query-based system to establish a user's relevance to a particular task, or another user—along the lines of the method in patent << >> with the efficient method to profile users.

The system of SDI can also be used as a confidential database for the purposes of generating statistics from sensitive data. For example, as a trusted system, manufacturers might be willing to provide information about their productivity, margins, retainment rates, production efficiencies, yields etc. The central SDI server could generate statistics, globally for the manufacturing sector, and then individually for each manufacturer—as it relates to the information provided by other companies. Similarly, it would be possible to use such a system to compare salaries across different universities. While an individual university might be reluctant to reveal information about its pay-scales to other universities, in the aggregate this information is not sensitive—and a survey on salary can be useful to both employers and job candidates. SDI is used to securely calculate statistics, without revealing any information that might compromise the privacy of a single employer.

Note: Describe secure virtual database involving matching by expertise and sharing of human resources (e.g. virtual work groups). Each member (a commercial interest) has a certain "resource sharing policy" which defines 1) what entities or types of entities s/he would share resources with. 2) If so, on a per-entity or per-entity type basis, what types of their resources (e.g. type of skilled employee and for what TYPE of outsourced task) would the entity share. It is an obvious extension to look at sharing of code, technology, intellectual property. A major challenge and limiting factor being how well informed SDI, the neutral'intermediary can be made aware of the needs/requirements of a company such that it can make evaluations entirely on its own regarding highly confidential materials with which it can accurately predict the basis for a deal WITHOUT disclosing to the prospective recipient what the technology or know-how entails (which could compromise the value of that asset should a deal not eventuate).

Further describe an extension/adjunct of the present human/technical resource sharing method whereby corporations (in particular high tech and Internet related) may employ SDI to utilize the above information regarding their human and technology sharing synergies in order to detect and recommend strategic (e.g. equity sharing, merger, acquisition etc.) relationship opportunities between the entities. B to b and even b to c iamworthit user behavior combined with text analysis should also provide revealing clues about what types of companies tend to share similar customers and provide similar (complementary or competitive) products and services which may suggest that such synergies are potentially available (Geographic region may also be detected or inferred). Again the disclosure of detailed business information is very helpful and a data release policy defining the parameters for such strategic initiatives may be critical in order to determine what companies may be potential candidates for which initial feelers (of high level information disclosure) would be appropriate to put out to a prospective company to determine mutual interest and/or further basis for expected synergies.

Note: Reference FAQ routing scheme in co-pending patent application, also add a section on using customized prices and promotions scheme to assign a price to a task, query response or virtual work group participation (where potential synergies are identified between the parties).

13.3 Virtual Tags for Annotation/Information Filtering

In this extended application of SDI, we allow users and other third parties to annotate objects (physical and virtual) with meta-information, either to remind themselves about a previous interaction in the future—or as a system of "knowledge learning", where systems of users leave useful information for other users. Information is left in the environment, leaving a trail for other users.

For example, the information that is tagged to an object, referred to as a "virtual tag", can contain a pointer to other relevant information, such as a survey of a film by a third party, or the user's own comments/feedback. For example, a restaurant listing could be annotated with meta-information about the quality of the food and service. Such information, when provided by a wide sample of users, can provide robust information about objects. The information that is used by a particular user can be filtered—for example, weighting the opinion of a respected restaurant critic, or weighting the opinion of users with common profiles (when that information is available).

Virtual tags can be assigned to objects with physical locations, and the information triggered based on the physical location of a user (using LEIA technology). Virtual tags can be assigned with expiry dates or other time-sensitive information. An individual user might leave an "action item", for example—next time I return to this object (eg. web page/ vendor) be sure to perform this task, enter this query, check this link for new information. As another example,

after a conversation with an SDI-enabled user it is possible to tag that user with some notes, to remember the conversation the next time the two users meet.

The technical innovation that allows this use of virtual tags, in addition to the protection of privacy, is that we allow users to annotate information to objects that they do not directly own through a system that separates virtual tags from the content that is tagged. In particular, tags can be stored (either at the ISP-level proxy, or main SDI server) for associated web pages, and exchangedlretrieved automatically when the object is accessed. The virtual tags can be used in conjunction with target-object profiles that are generated through SDI for web pages (and approved by vendors).

Virtual tags can be searched, using relevant terms, locations, or times, and can also contain links to authorative information, such as audio and/or video.

Tags are encrypted, so that only SDI-enabled users can access them. Tags are also associated with the pseudonymous ID of the user that left the information (although they can be anonymous, an associated profile allows more accurate collaborative filtering techniques). Finally, users can leave data-disclosure policies, embedded into tags—to certify the properties of other users necessary to release the information. When tags automatically are time-stamped with location, and time, and other information we allow for this information to be "fuzzed", as disclosed in the section on Randomized Aggregates, to protect a user's identity.

In the physical world, implementation of meta-information in a user's physical information, can be viewed via head-up displays, video cam monitors, wearable computing devices, or audio pieces. The information itself can be embedded directly on physical objects, for example on magnetic strips or via. visual encoding techniques--or the appropriate information can be accessed from a secure remote database based on the user's physical location (using LEIA location technology); or bar-codes that provide a universal identifier for an object.

13.4 Information Personalization Without Vendor cooperation

We can also rely purely on information provided (as virtual tags) by users about a web page, or other physical or virtual object, to filter and personalize information that is displayed and recommended to a new user. For example, because virtual tags are stored and accessed independently of the source of the object (e.g. not on the vendors own web server), they can be automatically picked up and associated with a page, even without a vendor that is subscribed to SDI.

The user's client-level proxy server can annotate web pages with relevant information as they are generated, for example adding mark-up to relevant information, or suggesting a link that the user should read, with notes that may be useful. Similarly, the database of tags can be used for an advanced web-search.

The ability to personalizedlprovide recommendations to a user, can be used in conjunction with a cache engine, that might be located on an ISP server. For users that follow advice, the ranking of recommendations is correlated to the information that the user reads, and therefore pre-caching can be more accurate, because the information about user-recommendations can be made available to the cache engine. Similarly, this is useful for advanced fetching to a the client-level SDI proxy, for example during idle time. Virtual tags associated with web pages are encoded to: (1) prevent a non-affiliated cache engine from using the tags; (2) prevent a competitive infomediary service from taking advantage of the "community information".

13.5 Demonstrating Value to Vendors

Vendors that subscribe to SDI can provide more attractive offerslproducts to users, based on information about the wider activities/interests of a user, on other vendor pages, and in the physical world (of course, only to the extent that this information is authorized by the user). Vendors can concatenate information from the client-level proxy, the vendor-level proxy, and the ISP-level proxy. In particular, wider information is vital for a good first-introduction to a user, so that information and products can be made relevant from the start. Similarly, such information (e.g. the user likes this type of music, but has purchased this CD) is very valuable for vendors that sell "content-based" products, for example books, and CDs. Vendors can personalize their service with, for example, a recommender system that interoperates with profile information furnished by SDI.

We can demonstrate this value experimentally, for example we can offer a vendor a free-trial and present personalized information/advertisements to one group of SDI users, and regular advertisements etc. to another group. The increase in vendor revenue can be estimated from client-level monitoring of the change in purchase volume achieved with well-focused solicitations *on the vendor's own business*.

Also, we can monitor the performance of vendors that use SDI technology, and provide measurements/metrics for new and prospective vendors (for example on click-through rates and transaction rates) etc. All of this is possible because the system of SDI has access to a user's client-side proxy machine, and the vendors themselves cannot block this collection of data.

NOTE to David -If you agree, it may be clearer o define "infomediary" in conjunction with the iamworthit section as a user-centric SDI which typically incorporates the iamworthit and/or community dollars models and go through and replace iainworthit elsewhere in the spec of "infomediary" use in the title "User "infomediary

14. User Infomediary: The IamWorthit Model

In this section we describe a key extension of the user-centric SDI model, that provides users with an additional incentive to provide information to vendors/advertising networks, in addition to the benefits from receiving well-targeted information and products. We introduce a new currency, termed "community dollars", that allows vendors/advertisers to compensate users for providing information—but tie the compensation to users making a purchase, so that: (1) users are incentivized to provide information that allows vendors to push relevant advertisementslproducts; (2) users will also be more likely to make purchases at a site for which they can receive discounts via community dollars; (3) providing users with community dollars will increase the number of hits to a site.

14.1 Time-of-purchase Competition

When a user specifies this "time-of-purchase" competition option to her SDI client proxy, SDI can automatically provide competitors with information about a user's product or service requirements, and a user's profile, before a user makes a purchase. This will facilitate competition between vendors, and can lead to better prices and offers for users. We allow vendors to opt-out of this scheme, and prevent the system of SDI from informing competitors about offerslpurchase-requests. Of course, in this case their user can decide not to visit such sites. The system of time-of-purchase competition enhances vendor competition, and can also help to reduce the costs of entry into a market, because name-recognition becomes less important. New vendors can simply register with the "purchase referral" service, and cherry pick the products that they specialize in.

This is a "next-generation" e-commerce service. Current shop-bots, for example "Junglee" at Amazon.com, www.shoptheweb.amazon.com, provides a static comparison shopping service, using static price information. A user can specify a product, and receive price information about the product from different suppliers. However, there is no dynamic competition on price or features. The buyer driven service for flights offered by www.priceline.com is more dynamic, in that a seller is found to match the price that a buyer bids, but does not promote competition between sellers. In fact the sellers can make excess profits from the pricing errors made by users.

We can use profile information, and historical transaction information for similar transactions, together with the customer price/promotion algorithm disclosed in co-pending patent << >> to negotiate on a deal with a vendor that will optimize the value to the user. Profiling of vendors, and user transactions, can allow users to avoid making bids that are too high and losing value (airlines in priceline.com can profit from inaccurate user bids).

We enable vendors with competing products or services to receive automatic notification when a user is interested in the purchase of a particular product. A vendors can also receive information on the profile of a user, and the offers made by other vendors; and submit counter-offers to a user via the user's SDI-enabled client. The user can then be presented with a final set of offers, before making a purchase decision. In a variation, the user may actually initiate the processing submitting an initial offer to a vendor or a collection of vendors for an item, shelhe is interested in (whether in a traditional store-front or even an on-line auction house). IamWorthIt protects the value of a user's profile through

competition, because instead of using a profile to practice price discrimination and extract excess profits from a user, a vendor must use a profile to offer the most appropriate product to the user at a competitive price.

The IamWorthIt model provides an environment in which a market is established that is dictated by users by facilitating competition by/between vendors vis-a-vie the release of information pertaining to offers presented to users of a disclosed profile (or more typically pseudonymous user profiles) by vendors to competitive vendors selling the same products or services. In addition to enabling a more competitive market for electronic commerce, the IamWorthIt model also increases the value of user information, that the user can choose to disclose to vendors to enhance customer targeted promotions and targeted pricing and other on-line marketing strategies, and also to withhold from vendors information in order to elicit optimal offers.

For example, the information that a user discloses to a vendor could include its sensitivity to discount offers, customer loyalty with other vendors, value responsiveness, (bargain driven), high quantity purchases (for only those categories which the user makes frequent or large purchases).

When a user selects the SDI time-of-purchase competition option the client software monitors the user's web actions for queries and browsing related to purchases. The client identifies other vendors with similar products or services, either using a static "web index" that maintains vendors in particular product domains, or through dynamic profile matches between the target object profile of the web site that the user is currently browsing and target object profiles of the web sites of other SDI-enabled vendors.

Purchasables which are similar or the same to another purchasable on other vendor's web sites must be identified in order to determine which notification of a browsing action or purchasing request by the user is of relevant interest to those vendors. This may be achieved by solely relying upon meta data embedded in the pages, or with static information that classifies vendors into product-domain categories, such as a Web directory. Key terms and other attributes associated with the items, automatic classification and clustering techniques applying usage statistics and content (associative, numeric and textual attributes as described on the parent issued patent) may be further deployed as additional techniques for purposes of identifying similarity at the level of the target objects. Classification and clustering techniques can be deployed to identify similarity between vendors at the level of target objects.

Vendors are notified, and provided with the ability to access the profile of the user, either with client-level processing or through the release of an anonymous profile to the vendor. Vendors typically will wish to construct offers through a rule-based engine, data-mining techniques, or automatic collaborative filtering techniques, as disclosed in co-pending patent application "System for Automatic Determination of Customized Prices and Promotions" and U.S. Patent #5,754,939, "System for Generation of User Profiles for a System for Customized Electronic Identification of Desirable Objects" as such techniques may be deployed by the vendor directly or via the Secure Data Interchange representing the interests of the Vendors.

User profile information may include a temporal profile of the user's present activities, including search terms, recent page navigations, what pages is the user observing presently (and the profile of this page) or even his/her present physical location as well as the general user profile (any portion of the above particularly the latter may of course be withheld from the vendor). In the preferred implementation, vendors are also provided with a (client or web-based) rules interface which enables the vendors to input pre-stated rules with which the system may solicit and respond to competitive offers automatically. If pre-stated rules are used to automatically respond to a notification with a competitive offer, the nature and degree of discount is typically determined in accordance with the nature and degree of the original or previous offer and/or the user profile as disclosed by the client-level proxy/server to that vendor. In lieu of manually entered rules, co-pending patent application entitled "System for Customized Prices and Promotions" or another similar algorithmic methodology may be used as an aid by the vendor in order to automatically determine a competitive offer (or subsequent responses thereto). These techniques can also be used in conjunction with a data-mining interface, in which predictive metrics as to selection, price and promotional type, may be determined in relation to the individual user or specific relevant user profile attributes for example, in accordance with a data analysis expert of the vendor (or representing the vendor via SDI) analyzing randomized versions of user profiles and randomized aggregate statistics.

A variation of the system is further disclosed which was intended (according to the co-pending specification) as an electronic assistant to telemarketers and other sales persons to determine offers and counter-offers which are automatically generated in response to (for example) rejections of the previous offer as well as counter offers by the user. This dynamic system was originally designed forsalespersons to optimize the expected profit from each

customer (in view of the general user profile and the offer user responses up to that point in the negotiation). As such, this technique could be readily extended to the current application in which the previous offers up to that point may instead originate from other vendors (instead of a single one), thus the system responses may be affected by the user profile as well as the offer response pairs up to that point in the negotiation process.

It is likely that vendors will not compete on price alone, but rather through added-value services such as offering loyalty bonuses, cross-sells, and two-for-one offer and added features. Vendors will choose this mode of selling to prevent simple price-comparison at the client conversely vendors may attempt to eliminate the features in order to create the perception of a better deal through marginal price reductions. Therefore the client will receive offers from multiple vendors, and after initial filtering of dominated offers, present a choice set to the user. Furthermore, we can allow vendors to offer payment to a client in return for displaying an offer to the user, and vendors can also bid for space on the user's web portal which is often represented as a profile associated with a pseudonym conjunction with a description of the ad space. The purchasing decisions of the user may be performed by an electronic representative of the user's wishes (as "user agent") implementing the techniques of pricing/promotion selection algorithms completely autonomously on behalf of the user. However, the best offer can only be presented to a user to the extent that the SDI client level software understands a user's model of "value", and can make appropriate tradeoffs between product features and price (as implicitly inferred by the system through the above suggested techniques or explicitly stated by the user in advance). Nonetheless, this is a hard problem, and we expect that the user will often need to make a final product choice decision.

The collaborative filtering techniques described in patent "System for Automatic Determination of Customized Objects and Promotions", can allow a user's client-level proxy server, termed the user agent in this section, to automatically analyze offers. The system can also be used to send initial offers to vendors, on the basis of historical information about the transactions that have been performed between other users and the vendor. Offers can (of course) be sent to a vendor and its competitors. Finally, after offers that are received from vendors are pre-screened, they can be automatically ranked for value—using a combined quality and price metric (again judged within a collaborative filtering framework). The goal is to leverage the database of other offers that have been accepted by users in the past, and form a model of vendors, to determine whether or not a user has received good offers. (i.e. we can exchange information within the system of Secure Data Interchange, and making more information available increases the efficiency of the market). Offers can be filtered and presented to a user in rank order.

Buyers might also form "buyer coalitions", on the basis of automatically detected synergies between their requests. This can give buyers more leverage in negotiation with a vendor.

If the user so desires, the client-level proxy can also automatically notify these vendors if/when a particular offer is about to be accepted by the user. For example, a time delay response in the client-level proxy actually processing the order requests could allow vendors a final opportunity to present another competitive offer to the user. In another less optimal variation, vendors are notified only upon the user agreeing to accept an initial offer received.

As an additional service to users the SDI-level proxy server can perform analysis on the offers that a user receives, through comparison with offers that have been received by other user with the best offer that has been received by any user for the same product, and with the typical offer received by a user with a similar profile to the user. This can be useful to a user because it will allow the user to reject all offers if they are non-competitive. The SDI-level proxy could also automatically identify for users the profile attributes that promote good offers, and the profile attributes that promote bad offers, as an informational service to enable users to gain better offers in the future, either through only revealing certain information or changing behavior to attain favorable profiles.

As an additional service to vendors SDI can provide enhanced profile information, aggregated from other vendors, to enable vendors to provide better focused offers than can be provided on the basis of the profile information directly associated with the pseudonym of a user. Certain portions of the user profile data that is unavailable for direct collection by the vendor (such as information that is collected on other sites including, in particular, competitive vendor sites) may reveal important information which enables the vendor to better target that user. As such the secure data interchange representing the collective users may aggregate, analyze and sell this data to the vendor.

14.2 Community Dollars

We allow users to receive compensation for providing personal data to vendors, information that has value to vendors because it allows information to be focused (for example relevant ads can be displayed to a user, based on his/her profile). The system of iamwothit credits users for information, and provides users with direct incentives to reveal profile information to vendors.

A vendor can sign up with iamworthit.com and agree to provide only the most restrictive type of community dollars, that can be spent at that vendors site.

Community dollars are the currency that vendors provide in return for the right to provide focused information to users. Dollars can be general (e.g. for a network of vendors), or very tightly focused (e.g. for a particular product, at a particular time). The user-centric infomediary acts as a broker, matching users and vendors. Another key role of the infomediary (e.g. the portal) is to protect the user from information saturation by controlling the flow of solicitations. (i.e. restrict the number of ads. that a user secs)

We can use meta-tags to restrict the way that community dollars can be spent. The tag associates the dollar, but the dollar is released within the system of blinded signatures (Chaum) so that a user that collects dollars over many transactions with different vendors can spend the dollars without compromising hislher private information about pseudonyms.

** DP. Need to disclose a system for aggregating dollars on the client-level proxy, for dollars collected under different pseudonyms.

Dollars can be restricted to a number of vendors, and also restricted in additional ways—i.e. they can only be sent if the user visits the site through a particular portal, cannot be redeemed at a competitor, are worth a bonus if redeemed with certain vendors, etc.

14.3 The iamworthit community advertising dollars model

The primary objective of the IamWorthIt model is to create a market for information about users. We allow vendors to pay in "community dollars" for adverts, dollars that can only be spent at that vendor (with the host site of the advert receiving a share of the profits). This provides vendors with the ability to gain long-term customers. Furthermore, so long as the user agrees to receive advertising from hislher iamworthit subscription offer, community dollars can be replenished at the rate at which advertisers are willing to pay for impressions. This provides users with an incentive to spend at the vendor's site, because the vendor can monitor (pseudonymously) the user's that are sensitive to discounts and other special offers (that are delivered as community dollars).

** DP. Be sure to disclose the method that vendors can track the user that receives community dollars without compromising a user's pseudonymity.

The main mode of the community dollars advertising model allows vendors to advertise for free, but provide community dollars to users, that can be spent at some later time. The cost of advertising can be linked to the success of advertising. Moreover, the vendor can direct offers and adverts to particular user profiles. The hosting web page receives a share of the vendor's revenue that comes from transactions involving community dollars. The dollars can represent "stored value", such as bonus points, that can be applied to special discounts for offers which are delivered via digital coupons and/or as "straight value" which could be converted directly to purchases thus are equivalent to real dollars at the point of transaction.

The community dollars can be "credits" that can be redeemed as real cash, credits towards discounts, and can be spent across a suite of sites, or limited to one site. The co-pending patent application entitled "System for the Automatic Determination of Customized Prices and Promotions" describes a comprehensive scheme which may be implemented in either on-line or off-line commerce environments. The system enables vendors to deliver a digital message in the form of a promise to a user (typically on encrypted form for purposes of targeting a user specifically). This promise is typically a discount for a product, set of products (or all products in stock) or may even include entitlement to special privileges for that user, thus it is termed a "digital coupon". The community dollars can represent special discounts for a user.

The user receives a financial incentive for receiving well-targeted solicitations, while preserving user privacy within the SDI system. The vendors support the community dollars through advertising revenues and increased sales volume. We

can also provide the vendor through which the user first subscribes a special "first screen" right, that allows the vendor to provide a user with his/her first impression as soon as s/he logs on.

In one variation all community dollars collected by a user must be spent back at the vendor site at which they originally subscribed (and also the site that hosts the adverts of other vendors). A user can spend the dollars with any vendors that are site partners of the original site. This provides the vendor an incentive to accept and promote the community dollars concept.

NOTE: DAVID Since the community dollars concept is unquestionably the most critically important piece of iReactor technology I want to be ABSOLUTELY POSITIVE it is comprehensibly disclosed technically.

The value of providing a user with targeted solicitations is estimated at approximately $300 to $500 per year (based upon $120 per 1000 targeted impressions at approximately 25 impressions per day). Given these significant benefits, a vendor can provide a user with a significant discount (in the form of community dollars). Vendors benefit from increased sales volumes.

When the price of items is less than the value of dollars, the vendor can limit the amount of discount that is available on any single product, or only allow community dollars to be applied towards customer discounts (which may nevertheless be quite substantial).

14.4 Preferred Implementation

In the preferred implementation we use an "electronic cash" infrastructure for the community dollar system. A user's SDI-enabled client-level proxy stores dollars that the user receives securely. Dollars are anonymous and non-traceable, so that the user can maintain a single "bank" of dollars, and aggregate dollars collected across pseudonyms for a single purchase, so long as the purchase satisfies the constraints on the dollars. Each dollar is created using Chaum's blinded signature technique, and also signed with the conditions on its use.

** DP. Need to disclose a technique to allow transfer of dollars across a user's pseudonyms.

This scheme allows vendors to monitor the offers that users respond to, because when a user presents a community dollar—the dollar can be validated to indicate the type of discount that it is, even if the identity of the dollar (i.e the serial number) is untraceable. SDI provides vendors with guarantees that users have once-in-a-lifetime pseudonyms, so redeeming a voucher of a particular type that is redeemable only at vendor V and was issued by vendor V allows vendor V to be sure that the voucher was issued under the same pseudonym, and has not been transferred to another of the user's pseudonyms.

In an alternative architecture, the ISP-level SDI proxy, or the web-host for the advertising service, can maintain community dollar "debit" accounts for each user. This is more limited, because it does not allow users to transfer dollars between pseudonyms without compromising privacy (revealing a portfolio of pseudonyms). However, in a scheme where advertisers require that agents have once-in-a-lifetime pseudonyms, and only release community dollars to be redeemed at their own site, this is not limiting. Both of these approaches are useful for "community dollar-enabling" numerous or all sites.

A vendor that allows community dollars to be spent does not need to implement a special community dollars/discounts program. The user can also be issued a special debit account dedicated to community dollars, that permits pseudonymous transactions without revealing a user's portfolio of pseudonyms.

** DP. disclose how community dollars can be integrated within a general privacy-protected SDI banking system. e.g. the ISP level SDI proxy allows a user to transact via the general SDI credit card account.

A portal site that hosts advertisers and users that subscribe to IamWorthIt can mandate that all community dollars are to be spent at sites that advertise on the portal site, and also only when the sites are accessed via the portal site. This technique will increase portal traffic. Portals can be expected to compete in terms of: (a) the fraction of advertising revenue that is turned over to users, in return for receiving profile information from users; (b) the level of advertising that users are exposed to; © the nature of the community dollars "package", i.e. what vendors can the dollars be used at etc. This can be useful to attract niche customers, that have common outlooks, interests, and business needs. The primary goal of the portal is to drive traffic through the portal.

** DP need a technical solution to ensure that $$s can only be redeemed if transactions are done via the portal.

14.5 Commercial Variations

(i) Vendor coalitions

Vendors may choose to form coalitions, to allow users to spend community dollars at any "partner" site. Vendors that have similar user bases can be automatically identified using collaborative filtering. (i.e. determining similarity with the present vendor, from the aggregate vendor preferences of a given vendor's subscribers). Also, these resulting metrics could incorporate predicted online spending by each user at each site. This could help to narrow the selection of sites the vendor wishes to partner with and/or the selection of these partner sites could be determined and presented to the user to even further narrow the selection for each user. All vendors in a coalition advertise, and provide cross-links and up-links to other vendors.

The coalition model is good for users, that are more likely to find products that they want. Vendors can share the risk of advertising, since dollars provided to one user by a particular vendor can be redeemed at another vendor. Advertising and community dollars increases sales volume at all vendors in the coalition.

We can even allow vendors to form dynamic and virtual coalitions within SDI, with a potentially unique coalition of vendors for each user. The coalition may consist of an optimal pool of vendors, as determined by SDI collaborative filtering techniques. The goal in this model is to provide users with a particular "brand" of community dollars.

We can allow each vendor to retain an exclusive right to advertise to each user; and also develop a portal for the coalition—that gives advertising prominence to coalition members. Portals will be expected to aggressively promote community dollars. Users that collect community dollars become loyal return visitors to the portal and its associated vendors. In the case the vendors do not generate the same value we can provide community dollars in proportion to the value that a vendor contributes to a coalition.

We can also provide targeted advertisements for the vendors at the portal, using the user profile to focus ads. The categories and links at a portal (that might include a search engine) can be re-prioritized (highlighted and/or re-ranked) in accordance with the user's preferences (as described above), and to favor subscribing vendors.

Vendors pay the portal site to advertise, and the portal provides community dollars to users in return for privacy-protected profile information. This model does not provide incentives for the portal to provide well-targeted adverts, because there is no direct link between a portal's revenue stream and the vendors' sales volumes.

A portal with community dollars that can only be spent under a single pseudonym at its partner sites also provides an incentive to users to interact under a single pseudonym—which in turn allows a portal to profile users across its complete vendor partner network. Users will access many sites with the same pseudonym. The system of SDI allows vendors to leverage the shared profile information as users browse web pages and products.

We can also lock users into a single portal—and a single coalition of vendors—with community dollars that "decay" over time, and must be continually replenished. In this way a user cannot pick and choose different portals, and different community dollars, but can benefit mainly from high web-browsing volume through a single portal. The value to vendors in terms of consumer lock-in can be considerable.

For example, a coalition of vendors can join to allow a user unlimited access over all affiliate vendors. The program can be sold through the existing marketing channels of each vendor, as well as through a portal directory of sites for those vendors. Vendors that join can be required to promote the program through their own marketing channels. Additionally, perhaps vendors are selected to cover exclusive physical regions (e.g. in the case of a set of ski resorts), or exclusive product categories (e.g. in the case of on-line vendors). Vendors can provide a community dollar-for-real dollar exchange, in return for becoming part of a vendor network. Alternatively, perhaps vendors provide an up-front fee, that can be recovered via dollars spent by users at their own site. Each vendor is obligated to sell the partner network community dollars, but is not necessarily required to promote the other community dollar vendors.

(ii) Enabling "Transaction-based" Revenue-sharing

In a second mode, the vendors provide discounts to users, and advertising is primarily "free". The only time that vendors incur a cost for providing a user with an impression is when they achieve a sale—because then the portal site receives a cut of the transaction price. The vendors provide users with community dollars directly. The dollars, which are stored at the portal site, allow user-spending to be tracked. This allows the portal to monitor when a sale occurs, not just a hit on a banner ad. With transaction-based revenue of this kind, personalization is critical. In this model the portal with give prominence to adverts from successful sites.

A portal site may forgo payment from a vendor in exchange for the increased click-through from a strong network of community-dollar enabled vendors. Value is credited directly to users for future redemption at that particular vendor's site. The community dollars provided to a user can be restricted, such that a user can only redeem dollars if the s/he maintains enough visits to the portal site.

Vendors can offer discounts on their own products directly, instead of providing the portal with money for advertising. The vendor only pays to the extent that its advertisements are well-targeted. The vendor could also request special ad priority. A vendor that presents advertisements to a user offers the user discounted promotional offers for products offered by partner vendors, in exchange for subscribing to iamworthit and receiving targeted impressions. These offers are in lieu of community dollars, and can be provided by partner vendors—maybe in exchange for a right to a number of ad deliveries for the vendor's own advertising purposes.

We can also require that users are automatically routed through a portal when accessing any partner vendor directly. The portal (and therefore the coalition of vendors) then receives exposure each time the user clicks on an ad (or link) to that vendor.

The portal may also provide benefits (e.g. additional advertising prominence) for sites which are responsible for driving traffic through the portal. Community dollars can be provided whenever the user accesses a site from the portal. Portals can offer free advertising to e-commerce sites (forgoing advertising fees). The portal provides discounts to users that purchase a product following a link provided at the portal.

A user receives the discount by validating a purchase with the portal, and the site agrees to provide the portal with a share of revenue whenever the user cashes in community dollars in this way (we do not rely on HTTP refer mechanism because that can be blocked and falsified. Furthermore, we do not rely on URL+extension correspondences, also not secure—instead rely on providing user's with incentives, and monitoring users that have earned community dollars.)

(iii) Delivering per-impression dollars

We can deliver community dollars on a per-impression basis, with vendors competing to offer users high values for being able to present an advert. The existing collaborative-filtering engine at a user's SDI client-level proxy can filter ads, and select appropriate offers, using community dollars as just another measure of the usefulness of a message. This is an alternative to providing dollars on a one-off (or even yearly) basis, for consumption via the vendor's site that the user subscribes to the service.

A hosting site can take a fraction of any dollars provided to a user. Alternatively, a site can convert the value into the community dollars to provide to the user, possibly at a preferable basis. The portal might also wish to convert its commission to credits for the user at any one of its partner vendors, with the stipulation that the user must access those sites via the portal in order to be able to redeem the credits.

14.6 Providing Loyalty Bonuses

We can use the client-side SDI proxy to provide vendors with "loyalty guarantees", that are credentials to verify that the user has executed no transactions with any competitor, under any of its pseudonyms. The client-side SDI proxy is in a unique position to be able to implement this monitoring, because no other system knows a user's portfolio of pseudonyms. The user can present its digital credential when visiting a vendor's site.

A vendor may wish to provide loyalty dollar credit; for example, it would be possible for vendors to offer user's credits if the user is a 100% loyal customer i.e. that she/he did not (over a specified period) do his/her purchases at the site of any competitor. For example, certain types of high value customers could be given considerable value in the form of credits or discounts as a result of demonstrated vendor loyalty.

A competitor could be determined by a simple known vendor list, e.g. through Yahoo or by matching similar products or service descriptions on sites across the web (e.g. using web crawler technology).

The credential can be time-stamped, to prove loyalty. It does not reveal any information about the user's other pseudonyms to a vendor, because many pseudonyms will exist that have not made any purchases from a competitor. Upon accessing the vendor's site, this credential may be presented to the vendor. One criteria for the above benefits could be that the user may visit a competitor site, and engage in interactions; however s/he should not transact with that vendor.

14.7 Off-line Variations

The community dollars may be stored on a portable smart card carried by the user, in which the community currency was originally loaded by the user's own computing device, or alternatively a kiosk or cash register controlled by the vendor or another third party whereby redemption may occur at any location where the subscribed vendor possesses a smart card reader. In another variation, the community dollars may be coded into a form which is bar code'reader enabled and distributed to the user electronically or potentially if used in conjunction with a traditional loyalty points program, additionally printed for the user at the vendor's physical location (such as point of sale or kiosk), applied in conjunction with purchases at the vendors physical location. At which point typically a new coupon is reprinted containing the updated secure information pertaining to the user's community dollar and/or loyalty points account.

In another variation, a promotion for a yearly allowance of community dollars could be printed as an advertising offer on a magazine coupon, newspaper insertion or direct mail piece which could contain a unique URL (typically the actual URL for the iamworthit community dollars subscription site with a unique post script as the character string ("code") identifying that particular vendor and/or that promotion) from which the user could subscribe to iamworthit, wherein the unique URL acts as an identifier for that particular vendor's promotional piece from which the user originally received the offer for his/her own community dollars promotion.

** DP. Careful, this could be used to breach a user's privacy, i.e. link a pseudonym and profile with a user's true identity.

Within SDI we could send physical solicitations to users, and allow users to access promotions pseudonymously. SDI can target a selected audience for each vendor.

** i.e. we need to user to be able to subscribe without a link made to real identity.

Example: An iamworthit card in accordance with the pseudonymous payment methods described above, such a card could be a direct extension of SDI into the off line environment. Users could use this card as an identifier such that when they travel physically from vendor to vendor, their profile data can be readily identified where data pertaining to their own behavior and policy (depending on their data release potentially part of the vendor's user profile data) is retrieved.

If a smart card is used this user profile data may not have to be remotely retrieved but may be stored on local memory on the card itself. In one novel variation, a card is done away with completely by virtue of revolutionary technological breakthroughs in being able to instantly and positively identify users biometrically using ISP's scanning techniques (which may in a variation be further combined with facial recognition techniques). Many vendors will wish to utilize user profile data in order to deliver promotions targeted discounts and promotions (see pending patent "System for Customized Prices and Promotions").

The co-pending application entitled "Location Enhanced Information Architecture" (LEIA) describes an integrated advertising delivery platform which selectively targets user personalized advertising based upon both the user's personal profile and the present location of the user which may suggest appropriate ads from vendors which are local to the user, wherein user identifiers (UID's) which could include any of the above identification media provide the essential elements for this user targeting platform.

For example, at a bookstore, we can recommend isles and particular books; at a supermarket, can play music preferences; smart-radio, play appropriate channels in a cab based on target object profiles (as meta-data). As suggested in issued patent "System for Broadcast of and access to Video and Other Data Using Customer Profiles" the appearance of relevant selections can be continuously scanned for, dynamically selected and presented to the user in the form of "Virtual radio station". Such a system can also be linked to a service for making an instant purchase, or linked

to a database (in conjunction with LEIA) to recommend where a user should physically go to make a purchase. For example, music selections that the user is presently listening to may be ordered.

Also, it is possible to provide advertising which is targeted to a user by automatically recognizing pre-existing commercials and replacing them with targeted counterparts. This can be done through the identification of previously played commercials for example, commercials that have been manually identified and classified. Upon recognition, targeted commercials (including those which are targeted by user location in accordance with LEIA) may be inserted into these spots, and delivered and/or pre-cached through cellular, satellite or radio communications.

At a public phone we can identify a user with hislher calling card, and deliver targeted advertisements, via the public telephone readout or delivering the targeted ads as audio messages in which server software at the phone switch (an ISP level proxy) recognizes if/when the user is put on hold and delivers audio and/or audio/video advertising to the user accordingly.

Targeted discounts and advertisements can be delivered at kiosks, for example using a credit card/smart card/other ID method (e.g. biometric...) Similarly, we can use credit cards to deliver targeted print advertisements on the backs of purchase receipts, e.g., supermarkets or fueling pumps or, alternatively, on a sheet dedicated for a advertising purposes conjunction with public copiers or printers or in another variation, on the cover sheet of incoming faxes which are sent to the user's fax machine or in which the user is otherwise identified automatically from the recipient's name field on the cover sheet.

Similarly, we can deliver targeted advertising and other information through cable TV systems, as desrcribed in the issued parent patent application entitled "Broadcast of and access to Video and other data users customer profiles", and co-pending application entitled "Broadcast & System for reduced memory terminals broadly address the use of cable systems as an interactive medium (in a bi-directional network architecture) for purposes of delivering targeted advertising targeted advertising and other information to the consumer based on user profiles". In this system customer behavioral data is collected at the digital set top and the upstream channel enables these profiles to be processed at the lead end server. These detailed profiles may then be subsequently transmitted down and stored at the level of the individual set top. The cable environment is a two way interactive medium. The bandwidth allocation is inherently asymmetric. Separate channels can push parallel adverts, which are selected at the set-top-box according to a user's profile. Each channel can have associated meta-data to allow matching at the set-top-box. As an alternative variation, full motion advertisements may be down loaded in the form of applets to the digital set top box and displayed to the user in similar fashion as described above.

14.8 Example: An on-line Gaming Site

Consider an on-line gaming site that has a network of affiliated vendors, that do not pay to advertise, but provide community dollars that can be spent either at the vendors--or at the casino. When users lose money at the casino the casino receives real dollars from vendors. The casino is one possible outlet for spending dollars—and a vendor only pays the casino (the host of its ads) if a user chooses to gamble on the site, and loses its dollars.

The gaming site becomes a portal, with links to partner vendors. Each vendor offers the user community dollars, that can only be spent back at that site or at the casino. However, the number of dollars which can be spent at the store is substantially less than the number of community dollars which can be spent at the casino. IF the user accepts the credits, whenever the user accesses the URL to the store helshe is either automatically routed first to the casino portal or to the vendor site whereby a prominent banner is displayed which is displayed to that particular user from which the user can conveniently engage in a casino gaming session.

If the user loses a substantial amount of community dollars he/she may regain the lost credits by spending a specified amount (in real dollars) at a partner vendor. This provides a safety-net for users. The cost to the vendor is the cost of the dollars that the user lost at the casino, and the cost of replenishing the user's community dollars (which can be used for further gambling). However, the vendor makes a sale—so the vendor is happy so long as the dollar value is a reasonable discount for the sale.

Clearly, the casino gains substantially through the redemption of these community dollars. The vendors can make an agreement with the casino where they only compensate a fraction of community dollars. A percentage of a user's

profits at a casino are paid in community dollars, another percentage can be paid in real dollars. Perhaps the casino can also provide vendors with a revenue share.

14.9 Implementation Details

We allow for community dollars that are restricted to particular products, and customized for an individual user. The dollar-object can contain two parts: the first part is readable to the user, and indicates the nature and the amount of the discounts to which the credits can be applied. The second part of the message is encrypted, and accessible only to the vendor, and is signed by the vendor to prevent any form of alteration. The information can contain the dollar credit to the user, the terms and conditions of the community dollars, a dollar amount, the pseudonym ID of the user, an expiration date, the terms and conditions of discounts and special offers to which community dollars may be applied in combination with a partial cash transaction.

The vendor must check that it has not previously redeemed any piece of community currency with the same identifier; the identity of the user is correct; the date; the terms and conditions. Some vendors may allow redemption of the community dollars at other vendors' sites.

** DP. See the paper in 3rd USENIX 1998, by Doug Tygar for technical solution.

14.10. Privacy Enhanced Marketing Applications

There are a number of commercial applications for a system that allows vendors to contact privacy protected lists of users in the pseudonymous user data base, wherein the contact, interaction, and business relationship with the vendor occurs under terms of complete user pseudonymity. Or wherein the user remains pseudonymous in his/her search and access to qualified professional services. In accordance with the parent patent application the pseudonymous communication may be either email, real-time text communications, voice (such as the pseudonymous telephony or Internet telephony).

For example:

(i) Financial Advice and Financial Planning Services

Often users are quite sensitive about the confidentiality of the release of this type of information related to personal financial matters and particularly with certain matters (and perhaps in general) prefer that their financial advisors were unaware of their true identities. Similarly, investment advice or sales communications by stock brokers are another application where similar user information is typically disclosed.

(ii) Insurance Agents & Brokers

For many types of insurance, (e.g. health, life, casualty) personally sensitive information is disclosed by users to their agents and brokers. Initially, before insurance services are purchased, it is possible that useful detailed quotes and/or insurance advice could be provided to a user pseudonymously.

(iii) Legal advisors

There are a variety of legal disciplines in which the associated legal services delve into highly sensitive personal information (e. g., bankruptcy law, divorce law, criminal law, etc.) Many lawyers also offer to first-time prospective clients a free consult in which such a privacy-enhanced communications system could be initially beneficial to the parties.

(iv) Family Counseling and Psychological Counseling

The parent patent application also suggests these applications which often involve the exchange of highly confidential personal information.

14.11 Business Models Leveraging Community Dollars

Free or discounted retail products with "niche" partners in each category Free dial-up ISP (as an independent ISP or a service to jointly promote free access with ISPs)
Free Cable and ISP service, Free pay-per-view (note that viewing patterns and the associated content could provide additional valuable user profile information)
Free phone service (e.g. advertise subscription service on screen phones or audio ads from pay phones)
Free prepaid calling card
Free print media subscriptions (magazines, newspapers)
Free book clubs
Offer any combination of the above with "deep discounts" for each (this can involve $350 community dollars per user or it may simply involve certain purchasing limitations per customer). Each vertical niche partner in exchange gets exclusivity within their own respective niches to target advertise to those users (e.g. retailers).
Free access to sporting events.
Free credit for casinos
Free lottery tickets
Free charity donations
Discounted hotel lodging
Monetary credit to a credit or debit card (either an iamworthit branded card or provided as a partnership with the card companies.
Monetary credit to a diner's club
Free subscriptions plus credit to retail buyer's clubs (on-line or off-line)
Credit or discounts for book clubs
Free musical concerts, or theater presentations, movies or
access to arcade entertainment
Free access to amusement parts or theme parks
Free golf season passes
Free commission fees for stock trading
Free commission fees for travel booking (if implemented for on-line users would be less compelled to search for travel information on-line though go off-line to make their bookings).
Providing a substantial credit in a user's electronic wallet in exchange for their downloading the electronic wallet to their client.
Free personal home pages (which community dollars could subsidize a high quality site).

14.12 Iamwothit Marketing Strategies

A number of marketing strategies are worthy to note. These include the following:

1. Allow the ISP to promote free Internet access through traditional means of Web advertising (impressions on ad servers) Web advertising whereby this advertising is effectively traded, directly for (e.g. pop-up) advertising on iamworthit. Furthermore, the ad server would be able to recognize through the associated domain names, the users which are coming from a competitor ISP. So long as that ISP is not a partner of iamworthit, the associated user would be selectively targeted with an offer of this sort "free Internet access" by subscribing to iamworthit".

Smaller ISP's would be particularly compelled by such offers to their direct target prospects. This is because they are operating on a "thin margin". Furthermore, both they and their small regional counterparts would be particularly vulnerable to this type of advertising by regional competitors from the same geographical area, during specified period of months of initial usage of the service, the share of profit due iamworthit could instead be committed to purchase additional advertising for the Internet service provider (or the balance of this profit traded-out in the form of additional advertising through the ad server partner).

2. The ad server partner could further become an exclusive partner of iamworthit on the following commercial venture:

Relationships as established with on-line merchants and other e-commerce sites. The vendor actively promote an offer to their customers through both off-line media (using a URL unique to that vendor) and on-line advertising through the ad delivery partner. The offer may say (as an example),"receive three hundred dollars in purchasing credit at Books a Million in exchange for subscribing to iamworthit (or receive five hundred dollars worth of discount credits at Books a Million In addition, as a further benefit to the vendor iamworthit could trade its own

advertising impressions with impressions on the ad server for the current offer (in order to reach a different base of users which are not currently subscribed). A particularly compelling industry for this application is on line travel inasmuch as a plaguing problem to this industry is the fact that many users use travel sites as an informational resource on available travel deals and packages, but ultimately book their trips directly through a travel agent (thus cutting out the travel site). A three hundred dollar a year travel credit would be a compelling incentive to many users to modify their current travel booking habits. Affiliate networks are also an ideal channel for these types of promotions because affiliate sites agree to participate (typically) purely based upon the degree of the profit sharing opportunity (which would be significantly larger than most types of transaction - based affiliate advertising).

3. Marketing Network Concept to Sell iamworthit - sites which offer a community dollars promotion could, upon the users subscribing to iamworthit, additionally offer the user with a down-loadable client based software which provides a small promotion in conjunction with a link to iamworthit. Each time a recipient of the email subscribes to iamworthit, a percentage of the value of that customer is credited back to the user in the form of community dollars. Each subscriber resulting from the current subscriber's email (though reduced) provides an additional credit to the original subscriber in accordance with the marketing network business model. If the site originally delivering the promotion is not an e-commerce site, a percentage of advertising revenues resulting from the subscriber (and potentially all resulting subscribers) could be used. It could be applied in the form of iamworthit advertising (or exchanged) for advertising in an ad server.

4. Free Community-based Content (e.g. broadband over the Web) - As an alternative to the community dollars scheme, particularly if a site does not have a principle focus on e-commerce, an attractive proposition to Web-sites could be the creation of premium content which is free to iamworthit subscribers as it would be subsidized entirely by community dollars. Each iamworthit user would be granted free access privileges to the premium content on all sites which are part of the program. Some content may be purchased and/or reusable, other may be entirely site-specific and novel. This model would be particularly appropriate for community sites which are largely member-based (or for example ISP-member-based communities) where much of their value to members is based upon information and other content which it can provide. It is conceivable that all iamworthit-enabled community sites would enable free access to their content by all other iamworthit customers (though it is possible that they may be mutually restricted if members of competitor communities as desired by the community site). If an ISP service is not already provided, a virtual ISP service could additionally be offered at a substantially reduced price or possible free (depending upon the number of community dollars left over. One could imagine further extending this present network of free content to free content iamworthit subscribers for free access to fee-based television programming or VOD services. In as much as community sites and television channels are becoming different media for delivery of the same information as the number of channels increases, VOD becomes technically enabled, and, most imminently, full motion video can be delivered upon demand over the Web.

5. Free Access To Subscription and Fee-for-use Information of the Web - In addition to the aforementioned free community site content, it would be possible to further provide free and automatic access to fee-based information on the Web. These costs may be able to be covered by the model across all or most sites depending upon the usage characteristics of its users (e.g. assuming advertiser/community dollars payment to the sites are averaged across users in accordance with the consumption patterns of the average iamworthit user). The identity (pseudonymous) of the user would have to either be disclosed to the site via the proxy or a unique pass code (as required by the site) provided to the user could be automatically entered upon the user accessing the fee for use area requiring the code. A directory (portal) of these fee-based sites would be a useful adjunct to subscribers.

It would be possible to offer websites the ability to become Internet service providers where the interface to the ISP home page would essentially be heavily branded to that site or portal. Companies like GTE already offer a "Virtual ISP" service in which the content to the ISP home page is unique to the ISP while the network is provided by the virtual ISP service. This model would be particularly compelling for sites which are largely community oriented and have a potentially loyal customer base.Moreover, interestingly, many of these community sites are offering many of the services and capabilities that a full-blown ISP would offer from its home-page, e. g., a portal interface, links to high-quality content, chat/forums,e-commerce, commerce affiliate links, etc.

6. Bundling iamworhit links with hardware with a PC manufacturer - Many PC manufacturers are now recognizing e-commerce as a very important sales channel. The present model would involve the P.C. manufacturer bundling a link along with a promotion for iamworthit. The promotion would offer the user cash credit for the user. The PC

manufacturer would also receive exclusive advertising rights to target users whose browsing behavior profile qualifies them as a future sales prospect. E.g., instead of cash credit as the profile is identified, the offer could then become modified to offer free hardware or credit to their purchase. Because PCs are highly portable, the advertising targeting techniques described in LEIA could add substantial additional value to advertisers. A similar model could be used for manufacturers of PDAs.

7. Providing a service to iamworthit partner vendors whereby their users may be automatically matched with each other based upon the similarities of their user profiles. This may be done for purposes of introduction or for introducing users into on-line chat or discussion forums. This service may be provided as a virtual service wherein a menu of different forums and chats are displayed and accessible from each iamworthit member site. (the underlying methodology for which is described in co-pending patent application "Virtual Community Service for System for Customized Electronic Identification of Desirable Objects"). In accordance with this specification, a variation of the service involves the process for identifying individuals who most closely match a given category or target object. For example in the context of the present implementation a category or content, merchandise or a purchasable being specially promoted may be the focal point of a discussion forum or chat room, which is automatically organized by the Virtual Community agent. Accordingly, a portal (or in accordance with the present trend) a site with which a portal interface is integrated utilize the present techniques for generating virtual communities for each category or sub-category of content on the portal or for direct access into a forum or chat room which was automatically created around that particular site (as the target object used as the matching criterion). As described, the user may navigate a hierarchical menu of virtual communities which may be constructed automatically according to the methods described which involves communities assigned to category, sub-category, and association with corresponding sites. Ideally in this scheme the portal is actually a "virtual portal" which may be utilized in providing access to the communities across numerous sites (and/or ISP home-pages). Users may also be navigated (at the individual user level) which along with their pseudonymous user profile data is subject to their data release policies. In a variation of the above schemes, if there is geographical information which is associated and which is released in accordance with the above individuals and/pr communities (e.g., as may be occurring or scheduled to occur in physical space), LEIA may be employed as a primary (or additional) selection criteria for navigating the present information accordingly.

8. Creation of an iamworthit online multi-store retail site - Instead of selling the community dollars model to on-line vendors, an alternative approach would be to establish a retail presence in a (or potentially multiple) retail niches. The primary business model would be to leverage existing large iamworthit subscriber base (involving the other various types of commercial partners) in order to dedicate a certain percentage of the community dollars (e.g. thirty percent or approximately one hundred fifty dollars per customer) which could only be redeemed at that multi-store retail site (and/or the value of these dollars could be worth more at the retail site). In addition, in this model, the independent advertising initiative of iamworthit would be geared towards community dollar credit of that retail site. It should be noted that, because if other outside competition occurs to the basic iamworthit scheme to a substantial degree there will not be a compelling incentive for users to adopt a more restricted form of valve (as retail credits at a particular site), versus accepting the credit from a competitor in the form of cash. Thus this model could provide a viable means for attaining a leading position in one or more on-line retail markets if this competition does not substantially exist.

9. Advertising in Exchange for Equity - A potentially attractive optional form of value, which could be provided to iamworthit customers involves equity shares in companies which advertise to the user (in lieu of community dollar credit or cash). This scheme is an ideal application for iamworthit in as much as iamworthit customers can be highly targeted and because many Internet-based start-ups are highly niche community oriented (thus iamworthit customers who are interested in the sites can be efficiently identified and targeted). Moreover advertising is typically very expensive which in the absence of accurate targeting may be of questionable value. It should be noted, however, that because the primary objective is to both find viable prospects and to engender an element of loyalty (which the equity model does). This scheme would be the preferred approach to advertising for sites which do not sell on-line where community dollars would be the preferred loyalty engendering scheme. In order for this model to substantially provide its desired advantages of increased advertising exposure to fledgling web based companies, the iamworthit subscriber base would have to be quite substantial.

10. Loyalty credits for off line retailers - Deliver through the back of sales receipts, kiosks or direct mail or on-line substantial purchase credit to retailers (e.g. grocers') customers, using the aforementioned technique of utilizing a unique URL to identify the vendor and/or promotion from which an iamworthit subscriber originally accessed the

iamworthit subscription site (thus identifying for both user and vendor the appropriate denomination and/or terms of community dollars issued to the user). In the preferred implementation, a loyalty card is used to identify the user thus enabling the community dollars value to be provided to the customer at check-out as straight credit or possibly an enhancement to loyalty credit. The user may also be identified via credit card or alternatively a voucher (or coupon) could be printed from the user's computer or from a kiosk which is typically situated near the entrance to the store and which could be activated upon insertion of a loyalty card credit card (or associated authorization code) and could also be used to disclose the user's community credit balance. A unique identifier for that voucher or coupon is provided and non-tamperability measures are provided such that the user's community dollars account can be appropriately debited upon redemption. Preferably, a pre-determined value is specified on each voucher (which could be predetermined by the service or the user) or alternatively, the total community dollars balance could be specified on the voucher along with the user's name/address and redeemable only upon presentation of valid user ID.

11. Auto Insurance Application - Co-pending patent application entitled "Applications for a Location Enhanced Information Architecture" describes a location-enhanced framework by which statistical methods are used in order to very efficiently and confidently extrapolate the most relevant attributes in predicting automobile accidents (or the avoidance thereof). The correlations from some of the existing metrics used may be refined using this technique e.g. LEIA is able to accurately determine the number of miles a user drives per week while the user will often lie about this, thus the basic model may be refined and more accurate information may be provided on a per-user basis. The scheme also enables completely new metrics to also be identified and utilized as well which may correlate the attributes location with time. It is conceivable that if a user provides access to this location-enhanced information by an insurer, that the insurer could in turn offer premiums, discounts or deliver credit to the user which could be added to monetary credit the user receives for personal information from iamworthit, for example, an iamworthit implementation which uses LEIA to profile and target users with ads by their location (e.g. while riding in an automobile).

15. Secure Data Interchange: General Examples

15.1 Assessing the Value of Data

- Plug together sets of data, and measure predictive accuracy.

15.2 Matching Data Across Vendors

- find patterns in common pseudonyms, denoting common areas of interest.
- use catalogues of order codes and item description to find similarities across data sets

15.3 Targeted Recommendations

- describe CDNow style project : similar customers defined by nearest neighbor and clusters; these are used to create recommendations for an individual customer.

15.4 Leveraging Portal Data

- data from portal to leverage data needs for ISP

15.5 Automated Generation of Customized Web Pages

Analyze customers for broad preferences in choice of web pages visited (corporate, Star Trek fan, etc.). This defines the initial look and feel for the page that greets them at their portal (a teen might enjoy lots of bright colors and sound clips, an investor would prefer a more staid design); different "skins" could be created to match the major categories of customers, and would designate both the graphical design and modules available on the page (e.g., a working stock-ticker for an investor, a real-time weather map for a jogger).

The web pages and information most frequently accessed by a customer would be given priority, and a hierarchy of usage could be developed. Since stock prices are of the highest importance to an investor, a ticker reflecting his portfolio value would stream across the top of the page. However, although he enjoys spending his profits on vacations and automobiles, these are only of secondary interest to him (as revealed by his on-line behavior), and so are relegated to a sub-menu on his web-page. As his usage changes, the priority level assigned to the modules would change as well, so that when a jogger purchases a treadmill for indoor running, his weather reports won't dominate the top-level screen.

Small children could have simplified browsers, with extra-big buttons and access to pages pre-screened by a "web-nanny" service.

SDI would be used in the initial phases to group customers into general categories based on their patterns of their web surfing, and would be used in later phases to adjust the content and style of their portal home-pages (based on what similar customers seem to be enjoying).

<<Mention fact that broadvision is trying to do some of these things .broadvision-like: (develop model for automated generation of customized web pages)

15.6 Analyzing Affinities

- suppose a vendor has a list of customers, and knows to some degree what web pages they visited after leaving vendor site. A large collection of customers taken from an ISP will contain their web-surfing behavior. Cluster web sites and cluster customers, finding cluster-to-cluster interactions.
- use this information to classify vendor's customers; gives vendor an edge in knowing customers' tastes.

15.7 Personalization of information.

Personalizing information on-the-fly requires that a vendor has a data model, for example that clusters its current user-base according to what they are likely to be interested in. Notice that it is only with more information from outside of their domain that users can be clustered with respect to their product space. Technology One: Need to be able to update vendor's data models, re-cluster users, on the basis of wider information about the users, without revealing that information. i.e. just provide the results of the analysis to the vendor, on his current user-base. This will not violate privacy, and allow rapid personalization. Technology Two: Need rapid profiling of a new user, in a way that does not reveal personal information to the vendor, i.e. can we provide an algorithm that the vendor can run on cryptographically secured profiles that accompany a new user to enable that user to be categorized. Alternatively, the vendor would need to request a categorization for the new user from SDI.

15.8 Enable user's to find appropriate information/products.

This could be more contentious with vendors, for example if we never recommend particular services.

15.9 Ad networks.

We need to allow ad networks to show ads relevant to the current user on the page. Approach: each ad network has a set of ads that it is currently running on a page. Can off-line develop a decision tree to decide how to assign a user to an ad, given a profile. Again, we need the user to come with the profile, and the ad network to have a secure evaluation technique that can run from profile, get answer, without getting any data.

Virtual ad networks. Track users across multiple domains. Leverage the user database. Fine level control over ads that the user sees. The trusted secure data interchange can operate as an "ad network", allowing for the placement of well-focused banner ads to market goods that are relevant to users of a particular content site. Electronic banner ads provide the potential for one-to-one marketing, when the advertising agency has information about the user that has just hit a site, together with information about what the user is doing local to a site. For example a car manufacturer is able to place a focused advertisement to a user that has just performed a search for new cars in a search engine, to a user that is known to have a large family and a high disposable income.

There are two possible business models. Firstly, an Internet content provider could purchase access to information placed by vendors and users within the Secure Data Interchange database. This information may be "rented" for a period of time, and then whenever a user visits the site of the content provider (possibly through the pseudonymous proxy server), the provider can query the data interchange for information about the user. The Internet content provider sells well-directed advertisements to vendors. Secondly, the data interchange could sell or rent data to an advertising agency directly, providing information in real-time to enable the advertising agency to provide more focus in its banner ads for its clients. "Per-transaction" pricing is a very powerful pricing model that is enabled with on-line banner ads. It is simple to monitor the number of click-throughs that are received at a particular banner, in response to an advertisement. In the off-line world pricing must be based on the number of impressions, or worst still, the number of mailings sent and it is more critical to understand the expected value of a campaign up front.

The proxy server could also act as an "ad network" itself, and sell focused advertisements for vendors, and purchase ad-space on the sites of content providers. The on-line domain provides this unique opportunity for quick experimentation with advertising strategies in order to get feedback on the likely utility of untested approaches. The system can use a hierarchical cluster tree to identify the most revealing items in a dynamically responsive fashion such that the profiles of all of the selections can be generated with the most minimal amount of interactions with the user (see "Rapid Profiling" section in issued patent entitled "System & Method for Customized Electronic Identification of Desirable Objects). Thus a more robust statistical model across multiple vendors is established as a result of the user's click through response of these intelligently selected virtual banners as well as other pages which are subsequently navigated through once the remote site is accessed via the banner.

In the preferred approach rapid profiling not only dynamically identifies and presents items which are most revealing of the other items in the collection, it also selects the users whose profiles suggest the greatest familiarity with these items (i.e., potentially correlated items). Furthermore, if the system's objective is to find new users or users who may be interested in the present vendor's other products, products for which little is known, then it will match users who are least familiar with exemplar items. The idea is to reveal the most significant data about the user profile with respect to the present collection of items of interest. Finally, rapid profiling can use direct explicit queries to determine interest on an item(s) or to collect demographic data on a user.

The target object profiles of advertisements on the ad server are matched against the user profile in order to automatically present the most relevant recommendation(s). Typically, the client-side proxy requires the host-level proxy to disclose the target object profiles of the products/services sold by the vendor. This data is stored as meta-tags in XML form and is encrypted. This data can be very useful to the user in navigation, filtering and search activities in the future or in a variation the ISP - level proxy a party (a neutral server) could store these target object profiles and selectively disclose relevant pieces of them (e.g. genre cross-correlations) to vendors, which are considered according to the disclosing vendor's data disclosure policy acceptable to receive this data. These profiles are not accessible to the client-level proxy but may be disclosed only if there are restrictions within the vendor's data disclosure policy.

In another variation, if the data to be disclosed to the vendor is acceptable to the original vendor but she/he is untrusting of the vendor, the data is received by the host-level proxy (another neutral third party) instead of the vendor, thus providing the disclosing vendor with an additional level of security, assurance about the use of his/her data while enabling the users of such a site to access all of the merchandise or content in a completely personalized fashion. Thus these XML tags are stored in association with, but on a separate server from the actual HTML pages stored on the vendor's site. Additionally, these profiles are constantly updated by user profile data conveyed to the host-level server which operates in distributed fashion.

15.10 Personalization of links

Adding value via data collected about this user and other users. Not just surfing data, but wider data-purchasing behavior etc.. On-line commerce enables the dynamic personalization of retail for each user. A virtual shop floor can be arranged to match the predicted preferences of each user. A commercial Internet site can leverage two types of information in order to personalize its product. For a new user, that has never before visited the site, it is very advantageous for the site to already know about the preferences of that user in order to personalize the goods and services that it offers. The information provided at the secure data interchange, and gathered from the transactions of a user with another vendor, is vital for this type of personalization to first-time users. The second type of information that an Internet site can leverage is information that it has collected from previous interactions with the user, information that is collected locally to the site.

15.11 Hospital database

In one data application of the client level proxy server, the user profile includes medical data which is obtained from medical records. (such as from hospitals or physician's medical records or potentially that of a health insurer). Typically, various physician's offices and hospitals which a patient (hereinafter "user") has visited over the years contains separate portions of a user's overall medical history, thus these various sources may be combined upon the user's request by downloading this data to the client-level proxy (or preferably, the user enters into a contract with those organizations in which all medical data and updates thereof are downloaded by the organization and/or an "agent" to the organization which transmits a request which is digitally signed by the user at the client-level proxy server. The origin of the request (the user) is authenticated and may be processed by a human or another agent located at the organization's host computer.

Because of the highly sensitive nature of medical data, there are potential user privacy advantages in using randomized aggregates. For example, a user's age, medical history of specific relatives (particulars of which could be more generalized) genetic data, numeric values associated with various medical tests, results for which are a numeric value. This data may be of relevance to pharmaceutical companies, alternative medicine vendor and clinics insurance companies hospitals physicians, clinics and home health care providers, the latter three of which may wish to advertise to patient prospects and extend their medical practices. The privacy architecture herein provided is a critical component for enabling access to user data by these commercial entities.

15.12 Smart Web browsing

Definitions: exemplar - the profile of target object or (as pertinent to following description), user profile which is "most like" the profile of the cluster to which it belong, perhaps a median metric.

The Platform for Privacy Preferences (P3P) provides for the ability to utilize XML meta-tags for purposes of annotating Web pages with comments from previous visitors to those pages. It further suggests that users with an affinity or otherwise associated with a particular category (e.g. of expressed interest) may both identify themselves as well as select a category (e.g. which they are associated with) and observe the annotations of other users associated with that category. One of the divisional applications of the parent case "System for Customized Electronic Identification of Desirable Objects" relating to the automatic creation of virtual communities suggests that users may be automatically assigned to particular communities (e.g. chat groups, forums etc.) which may be either automatically generated and labeled or constructed and labeled manually.

Additionally, users could further be allowed to rate the annotations to the pages, which could be a means by which qualifications of users to provide annotations could be measured by content domain (i.e. cluster). Their comments (and particularly future comments) could then receive a priority position in the annotated comments available. Future comments from users with a poor rating history for a particular content cluster may be deleted. A persistent interface feature on the tool bar or side bar may provide for annotations to also be accessed by users selecting certain profile features of users as they browse from page to page. E.g. identify the comments of a news article about abortion by users who are self identified as advocates of the Women's Rights Movement, ultra conservative senior citizens, teen women or those with a strong interest in alternative medicine or the Catholic Church or identify the annotations relating to Ford Motor Company by general Motors (though in the case of competitor annotations the above description suggests means protective safeguards for the recipient of the annotation).

The parent case further suggests that users may actively provide ratings in a completely privacy protected manner according to various criteria of pages they browse. A reasonable extension would further include being able to observe how these ratings relate in accordance with various user profiles (or groups which users identify themselves in association with). E.g. overall quality, aesthetic appeal, interest in the content, value (if it is a purchasable) etc. Each of these criteria's overall ratings would vary in accordance with the particular user groups which is selected by the user. Accordingly users may submit as a query a user profile or user cluster profile a page rating criteria a combination thereof and receive a listing of the pages of highest relevance to the search criteria. These may include, for example, the exemplar user profiles of users who previously most visited the site or the exemplar user profiles of the most predominant clusters of users who previously visited the site. In addition, the page rating profiles may be also displayed as well as related links which are determined to be most relevant to that collective group of users e.g. as statistically estimated from the referral logs or explicitly identified as book marks by those users as being of particular relevance or similarity to the present page.

As suggested and described in the parent issued patent application, these clusters may be accessed through querying (search) filtering (where relevant pages and/or annote to relevant page are "pushed" to the user as they appear on the Web), and browsing which includes navigating a hierarchical menu of users who are classified according to their passive behavior patterns and/or ratings which have been actively submitted as well as automatic hyperlinking (or automatically linking the presently viewed page to its "nearest neighbor"). Based upon the above criteria as selected by the user, ratings and annotations may be viewed in addition to these techniques are preferably deployed in the context of the navigational techniques as taught in the parent case.

The above description also describes the use of a hierarchical menu through which groups of users may be identified by their profile features (wherein a profile feature could even be a rating criteria itself of for e.g. an opinion via a site survey). These features could also include a user selected rating criteria by a certain type or group of user or users sharing rating similarities. These features could in turn be used to either selectively filter-out content which falls outside of that criteria as the user navigates the information by a variety of means (e,g. it could be used to selectively filter items or even categories in the hierarchical menu) or identify if/when pages are otherwise encountered where these user rating features are present (or metric features are predominant), thus displaying this user statistical information in conjunction with the ratings statistics and/or associated annotations if desired. (NOTE: CLARITY ??)

Other criteria for observing ratings and annotations may be in accordance with those submitted by organizations. As above described, recall that the user may also use one or more of the organizational approval criteria to also bias or completely filter selections as the user proceeds to navigate the Web. These endorsements may apply to individuals which may either provide annotations, ratings, news group postings, or editorial content.

Vendors delivering targeted on-line advertisements may also be interested in the above information, in particular the profiles of the most exemplary groups of users and their affinity ("similarity") toward particular ads, purchased products or product or content categories—as measured against user attributes or exemplar user profiles of the most relevant clusters of users based upon passive behavior and/or active ratings. This information may, of course, may be displayed (or selectively displayed) with the pages for the product (or service) as meta-data in accordance with the vendor's wishes. For vendors, these techniques may of course be applied exclusively within the scope of the web-sites of the vendors concerned (and/or additionally to users). For example, the site-specific page view correlations (including time spent viewing each page) in accordance with the dominant clusters, Exemplar user profiles and attributes of those users are certainly of interest to vendors to which those sites belong as well as affiliate sites on which their advertisements and/or syndicated products are advertised and sold remotely.

15.13 Location Enhanced Delivery System Architecture (LEIA) Enhanced SDI

For an exemplary application of the Secure Data Interchange technology, consider its extension to co-pending patent, entitled "Location Enhanced Delivery System Architecture" (LEIA), we teach a method for matching information providers and information recipients that utilizes location information, in addition to static and dynamic profiling information. The method customizes the information that is displayed on a private or public information device to the real audience in the vicinity of the device, instead of a pzedicted audience. LEIA collects an extremely detailed and coinprehensive information set about the daily activities of a user, enabling enhancement of the user profile with

location information and temporal activity patterns. The co-pending LEIA patent suggests appropriate application environments, for example in a smart home, an office, on a mobile shopping device, and in an automobile. A LEIA-based system stores personal information on users.

We can extend LEIA by incorporation with the Secure Data Interchange system that we teach in this patent. SDI enables the user to receive the benefits of powerful and well-directed information, but within a system that respects his/her privacy requirements. The interchange acts as a secure data warehouse for users and information providers, enabling information providers to target users without revealing private information to the providers directly.

Co-pending patent, "Location Enhanced Information Delivery System Architecture" (LEIA) customizes information that is displayed to an information recipient based on object profiles and physical location of users. Presents the information most relevant to the REAL audience, not a predicted audience.

One application includes "Smart Home Intelligence", where methods are disclosed by which users' real-time behavior may be profiled through their movement throughout their home, and specific interactions with the various network enabled appliances throughout the home. Other inputs may include the user's speech patterns (using voice recognition and text analysis). It could for example, note the user's speech content patterns in real-time. Such information provides invaluable dues as to the user's present activities, mood and interest state and may be processed by the presently described algorithms tuned with location/time features typically using the assistance of human data analyst to identify the key features and correlations. (This information may also provide enhanced information pertinent to the user's general, static preferences as well).

Other extensions of this scheme are also considered e.g. within the context of the user's office, or automobile and pedestrian activities. This application may thus extended the usefulness of the iamworthit model to advertisers in being able to target users through the presently anticipated on-line media as well as networked appliances and in either case, based upon the relevant context of users' present activities and behavior (and from this potentially their inferred moods or mental states) within their homes and elsewhere.

15.14 Extended Example: Vacation Packages

A vacation package organizer decides to begin a large-scale marketing campaign to target those people who would be the most interested in joining a new Caribbean Cruise. Although the vendor has a database of current customers, it is interested both in increasing the number and suitability of its potential leads.

Interfacing with the secure data interchange with which it is a member, the organizer identifies several possible sources of supplemental data: a LEIA-based travel discussion group, an on-line bookstore, and a Caribbean restaurant. These are found both by browsing through the interchange's internal list of members, and by using SDI-based data analysis tools, used within the interchange to automatically identify entities sharing common characteristics.

The package organizer then contacts each of these entities through the interchange, and negotiates different data-sharing deals: the travel discussion group is willing to exchange full information for a large travel discount, the on-line book store is willing to reveal the pseudonyms of users who have bought travel books in exchange for a per-sale commission, and the restaurant is willing to sell its entire database for a flat fee (and will provide an aggregated data set as a sample).

The vacation package organizer now chooses fairly basic data-mining algorithms to identify the individuals with the greatest potential interest in a Caribbean vacation; however, the organizer does splurge on a new neural network approach developed by a small software company. On a per-sale commission, the Software Company is willing to loan the vacation package organizer use of its data mining code.

First, the organizer decides which data sets to use. The initial results on the restaurant's aggregated data aren't so good (its customers turn out to not be very affluent), so the organizer declines the purchase of the full data set. However, it does agree to the conditions asked by the travel discussion group and the on-line bookstore.

The data provided by the discussion group and on-line bookstore, being in a common format, are moved in a secure fashion to the interchange's processing area, and are acted upon by the data mining tools, which are also in a compatible format. As per the agreement, the interchange forwards discounted Caribbean cruise offers to the members of the discussion group, and forwards standard promotions to targeted individuals in the book store's customer list. A few of these individuals respond favorably; these electronic transfers of money and passed back through the

interchange, which slices off a commission for the book store before passing the accepted offers back to the tour organizer, who learns the identities of the customers and can now count them as part of its database.

This protocol specification could even be digitally signed by the "owner" of the data as proof of ownership of the data and its associated restrictions by the owner, i.e., effectively a "digital deed" which is both legal and untamperable by any other party and thus acts as a legally binding proof of ownership and terms/conditions dictating how that data can be used.

16. *Future Directions*

(i) Digital Ownership Deeds.

A digital-deed is a digitally signed credential issued by a trusted witness of a transaction. It can be useful to have a secure record of a large ticket item that is bought or sold in an on-line transaction. It is convenient to have this verified and maintained, automatically. Provide increased confidence for members of SDI, particularly those which may wish to have a record or "back-up" verification of large transactions e.g. within the business to business implementation of SDI or for real-estate transactions, investments in small entities, etc. Additional data may further be associated as part of the digital deed credential, including any restrictions, stipulations, (including time-based), warranties, insurance, etc. pertaining to that item. Credentials cannot be tampered with by some other party, secured using cryptographic techniques.

In the domain of "information goods", SDI can automatically add an electronic "water mark" to digital information (e.g. a video, a CD, or a book), that associates a license for use, and terms-of-use. In this way (because the water mark cannot be removed, but does not affect the use of the data), the legality of selling/duplicating pure information goods can be verified by inspection of the water mark. Metadata can be used to assign licenses for intellectual property, along with data regarding terms of licenses, assignments, filing and expiration dates, etc. The seller can impose, within this untamperable credential, certain terms and conditions of who the buyer may be.

Credentials can be used to restrict the ability of minors to purchase pornographic, or other unsuitable material; and can also be tagged to products that are sold, so that such material cannot be exchanged within SDI without the correct credentials.

Credentials can also be used to create a "resale market" for potentially any and all items a user purchases (which has a potential resale value). The user that buys the good can be an "advertiser" in a resale market. The user will want to prescribe controls over the personal information made available to interested parties in the resale market. A user becomes a potential seller upon having made a purchase for an item (which is documented and recorded with the digital deed).

She/he is automatically asked upon transaction (by software on the client-level or ISP-level proxy) whether she wishes to have hislher item listed as a potential purchasable. If no, shelhe is asked if/when at a later times shelhe may change hislher mind. A typical price range for that type of item is presented and a question is asked about an approximate price range shelhe would be interested in selling for (though a stated range is optional since the preferred variation involves the use of a bidding scheme). The user may wish to control identifying information about themselves, i.e. they may wish to remain anonymous or pseudonymous. The current seller can however be provided with assurances about the user, before entering into a commercial relationship.

Regarding prospective future buyers, so long as the future transaction occurs on line (or conceivably even if it occurs off-line), a user credential can be a required stipulation to ownership of a given product. For example, some vendors may wish to impose a restriction which states that any product which they sell may not be sold to a competitor (based on stated characteristics and/or an explicit list of competitive vendors), or this will prevent an on-line black market for age-restricted goods. Such goods, e.g. information goods, can be protected at source—and water-marked to prevent unidentifiable duplication. Similarly, users (e.g. owners of kittens) may have certain personal interests whereby items or personal or sentimental value which are sold should be owned exclusively by certain types of users.

Another example may be the sale of corporate assets, whereby the ability to impose certain restrictions on these assets may provide negotiating leverage at the present time for the seller. The seller may state that if a presently submitted offer under the stated terms is not accepted by X (future) date, the offer will be cancelled. The seller can credibly commit to this negotiation strategy, via software credentials. SDI acts as a trusted mediator.

A seller could also make a believable threat to place a penalty on all buyers, with except for on some commercial entities, through irrefutable statements. The excused entity could be a competitor to the prospective buyer. The seller can also commit to a time limit, such that the seller cannot reverse the time limit. In general, the ability to make irrefutable claims improves the efficiency of negotiation.

Within the present system SDI provides the framework by which appropriate buyers and sellers may be matched together. It also enables a methodology by which the buyer interests are protected through the use of matching of sellers offers to competitive vendors (using iamworthit).

(II) Monitor Privacy Violations

We can also monitor sites to determine privacy violations. A human analyst can review the stated privacy policies of a Website, and monitor information that is used about an individual, from the personalization that a user receives. If the practices are consistent with stated policies then a credential can be issued. Otherwise, it may be possible to passively observe the degree (e.g. frequency and nature) of the violation, and what is the particular privacy violation. This can be done in combination with any history and details of privacy litigation (e.g. damages). This information could be very useful to a privacy insurer in determining at any given time, what sites are insurable and the associated risks of insuring them.

This information may also be used to indicate to iamworthit users (e.g. statically via a "black list" or dynamically while browsing) what sites violate user privacy, what have been the nature and (potential) extent of violations to other users as delivered via the client-level or ISP-level proxy. The subscribers may choose to adjust their data disclosure policy settings to assume (for example) single site pseudonymity or complete anonymity whenever visiting those sites and only certain profile information (or none) may wish to be released accordingly. This information release benefits also iamworhit in as much as in the preferred implementation, it automatically provides its users with privacy insurance, paid from a share of ad revenues (unless the user electively opts out).

We can also provide reports to users, including specific information relating to their own interactions and transactions with a vendor. The information may include pageview statistics, time spent viewing the site or its pages by visitor or collectively, transactions and amounts transacted. In a variation, the service may provide user specific information regarding perks, benefits and bonuses (e.g. community dollars/discounts) which the user is entitled to.

** DP. Technical disclosure for above?

(III) Vendor-centric Tools

Finally, metrics are useful within electronic commerce—for example, how many hits does an advertisement receive, what is the profile of users that hit the advertisement, how well does targeted advertising perform, etc. This information is available within SDI, through client-level proxy monitoring, and can he released or aggregated, according to user data-release policies.

There are a variety of different auditing, validation and reporting services which can conceivably be performed and, furthermore, this data can be analyzed and converted into digital credentials which may be useful to sites. Other issued directly to the site for the benefit of validating the associated information to visitors to the site (these credentials could alternatively be displayed virtually via the ISP-level proxy).

Auditing and validating click-through rates of various on-line vendors and advertising services. Data collected by iamworthit at the ISP-level proxy server and by the vendor-centric SDI service at the host-level proxy may provide a form of a reporting service in which SDI could be used to perform on-line usage and transaction measurements across a variety of sites (including competitive sites personalization enabled and targeted ad delivery sites etc.)

Existing reporting tools are typically implemented on a vendor's own site, however, the present scheme which monitors the user at the click-level proxy is valuable for purposes of suggesting correlations of visitation and behavior relating to

sites external to the vendor with those simply providing raw aggregates statistics regarding click-through clusters and associated traffic volume on that particular site sometimes as a fraction of time.

The primary goal of SDI is to protect the privacy of users, while providing incentives for users to provide information (within a controlled environment), and allowing users to reveal personalization within a controlled environment. However, it is conceivable that vendors might try to break the system, and make links between users across different vendors. Relevant information could include: the timing of access to and from sites, the time between clicks, typing cadence, and cursor movement. The system of SDI disables the Netscape cookie mechanism, removes referral tags from HTTP messages, and anonymizes routing information — however if any one of these systems is not in place (perhaps a user is not subscribed to SDI) then vendors can clearly use all of these to gain additional tracking accuracy.

The Vendor-centric version of SDI (unlike the user-centric version i.e. iamworthit) may wish to make optimal use of information which is collected from an unidentified visitor's click stream in order to make inferred "guesses" about who the user might actually be (e.g. a user who is unidentified by cookies, digital certificates, their client-level proxy or their customer account or credit card number). The vendor-centric version of SDI may utilize the data it collects from users across multiple vendors in order to make certain inferences about the identity of that individual.

A user's overall profile (available to all vendors) may reveal clues about the identity of user, if it is not sufficiently randomized.

** With vendor cooperation, referral across sites can be used—in conjunction with timing clues. This can be very valuable information, in conjunction with general profile information, similar typing, click rates, interests etc. i.e. vendor cooperation can recreate some of the information that is lost when the HTTP referral log is carefully "washed" via the SDI proxy server.

A robust user profile, across multiple vendors, is very valuable. Such a profile can be very revealing of a user's identity, allowing a more complete profile of the user's interest. Many of the same techniques described in SDI to extrapolate user preferences (including data mining, collaborative filtering and text analysis) may uncover unique identifying features.

Additionally, the fact that this information may be correlated with other unique identifying features, such as dynamic manual features of click stream, (time between clicks), cursor movements and typing cadence may enable the construction of a rather complete picture of the user and his/her identity. (e.g. using neural nets..)

A user could also be presented with questions to answer, to establish identity--combining a certain amount of real data from a user's SDI proxy with the vendor's profiles of users.

This can help to positively identify a user across sites.

Multiple cooperating sites may use the present techniques to predict, both when a user is likely to have returned to his/her site, and when a user is likely to have accessed another vendor site. This can allow a vendor to update a user's profile (with cooperation with another vendor), on the basis of what a user does at the subsequent site. Profiles can be updated statistically to allow for uncertainty.

Cookies cannot be leveraged under the standard SDI model, because they are disabled.

Theoretically, all vendors could, cooperate in the present initiative (to identify and comprehensively profile the user without his/her consent) while not authorized SDI to do anything further with that data (for data sharing purposes). If each vendor logs the path of users (under pseudonyms) as they pass from site to site, it could be possible to use data-mining techniques (with timing patterns) to make statistical predictions about user pseudonym portfolios. However, data is likely to be noisy, and the amount of data that is necessary is huge, especially when many users may hit a vendor's site simultaneously.

17. *References*

[APPEL] "A P3P Preference Exchange Language (APPEL)." World Wide Web Consortium Working Draft (Work in progress), 1998. URL: http://www.w3.org/TR/WD-P3P-preferences

[CKR97] Connolly, D., R. Khare, and A. Rifkin. "The Evolution of Web Documents: The Ascent of XML." *World Wide Web Journal* 2 (Fall 1997): 119-128.

[IRDF] Lassila, O. "Introduction to RDF Metadata." W3C NOTE 1997-11-13. URL:http://www.w3.org/TR/NOTE-rdf-simple-intro

[KRR97] Khare, R., Rohit, and A. Rifkin. "X Marks the Spot: extensible Markup Language opens the door to a motherlode of automated Web applications." IEEE *Internet Computing* 1 (July/August 1997): 78-87

[KR97] Khare, R., and A. Rifkin. "Capturing the State of Distributed Systems with XML." *World Wide Web Journal* 2 (Fall 1997):207-218.

[MCF] "Meta Content Framework Using XML." Submitted to W3C, June 1997.

URL: http://www.w3.org/TIUNOTE-MCF-XML-970624

[MCFXML] Guha, R.V., and T. Bray. "Meta Content Framework using XML." Submitted toW3C, June 1997. URL: http://www.w3.org/TR/NOTE-MCF-XML/

[Mil98] Miller, E. "An Introduction to the Resource Description Framework." *D-Lib Magazine* May 1998.

[NS] Bray, T., D. Hollander, and A. Layman. "Name Spaces in XML." W3C Note, January 1998.

URL:http://www.w3.org/TR/1998/NOTE-xml-names-0119

[OPS] "Proposal for an Open Profiling Standard." Submitted to W3C, June 1997. URL:http://www.w3.org/TR/NOTE-OPS-FrameWork.html

[PRIVACY] "Privacy and Profiling on the Web." Submitted to W3C, June 1997.

URL: http://www.w3.org/TR/NOTE-Web-privacy.html

[SCHEMA] The W3C RDF Schema Working Group,

URL:http://www.w3.org/TR/WE-RDF-Schema/

[SGML] *information Processing—Text and Office Systems—Standard Generalized Markup Language (SGML),* International Organization for Standardization, Rdf No. ISO 8879:1986, 1986.

[SPEC] "Resource Description Framework (RDF) Model and Syntax."

URL:http://www.w3.org/RDF/Group/WD-rdf-syntax/

[URI1] *RFC 1738: Uniform Resource Locators (URL).* IETF RFC1738 IETF (Internet Engineering Task Force, ed. T. Berners-Lee, L. Masinter, and M. McCahill. 1994.

[URI2] *RFC 1808: Relative Uniform Resource Locators.* IETF RFC1808 IETF (Internet Engineering Task Force). ed. R. Fielding, 1995.

[VC] *vCard Home Page,* URL:http://www.imc.org/pdi

[W3C] *The World Wide Web Consortium Home Page,* URL:http://www.w3.org/

[XML1] Bray, T., J. Paoli, and C. M. Sperberg-McQueen. "Extensible Markup Language (XML): Part I. Syntax", World Wide Web Consortium Working Draft (Work in progress), August 1997. URL:http://www.w3.org/TR/WD-xml-lang.html

[XML2] The W3C XML Extensible Markup Language Working Group Home Page,

URL:http://www.w3.org/XML/